

**Clonado posicional del gen *VRN-D4* en trigo hexaploide  
(*Triticum aestivum*)**

*Tesis presentada para optar al título de  
Doctor de la Universidad de Buenos Aires, Área Ciencias Agropecuarias*

**Néstor Fabián Kippes**

Licenciado en Biotecnología, Universidad Nacional de San Martín, 2010

Lugar de trabajo: Universidad de California, Davis, CA, USA



**FAUBA**

Escuela para Graduados Ing. Agr. Alberto Soriano  
Facultad de Agronomía – Universidad de Buenos Aires



## COMITÉ CONSEJERO

Director de tesis

**Jorge Dubcovsky**

Lic. en Cs. Biológicas (Universidad de Buenos Aires)

Dr. en Cs. Biológicas (Universidad de Buenos Aires)

Co-director

**Gabriela Tranquilli**

Ing. Agr. (Universidad de Buenos Aires)

Dr. en Cs. Agropecuarias (Universidad de Buenos Aires)

Consejero de Estudios

**María Laura Appendino**

Ing. Agr. (Universidad de Buenos Aires)

M.Sc. Producción Vegetal (Universidad de Buenos Aires)

## JURADO DE TESIS

Director de tesis

**Gabriela Tranquilli**

Ing. Agr. (Universidad de Buenos Aires)

Dr. en Cs. Agropecuarias (Universidad de Buenos Aires)

JURADO

**Ana María Castro**

Ing. Agr. (Universidad Nacional de La Plata)

Dr. en Genética (Universidad de Córdoba, España)

JURADO

**Fernando Carrari**

Ing. Agr. (Universidad de Buenos Aires)

Dr. en Biología (Universidad de Buenos Aires)

JURADO

**Esteban Hopp**

Lic. en Cs. Biológicas (Universidad de Buenos Aires)

Dr. en Cs. Biológicas (Universidad de Buenos Aires)

Fecha de defensa de la tesis: 2 de DICIEMBRE de 2015

## AGRADECIMIENTOS

*Quiero agradecer especialmente a mi director el Dr. Jorge Dubcovsky, no solo por haberme dado la posibilidad de trabajar en este proyecto, sino también por ser un ejemplo de dedicación, compromiso y estímulo contante.*

*A mi co-Directora y mi Consejera de Estudios por haberme tenido paciencia y estar siempre disponibles. A todos mi compañeros del laboratorio de Genética de Trigo de la Universidad de California, Davis. En especial a Francine, Stephen, Tyson, Hans, Nico y Juan por la ayuda constante y el apoyo moral. A mis ayudantes, estudiantes de la Universidad de California Davis (Anna, Maisie, Karla y Marielle) porque sin ellos hubiera sido imposible llevar a cabo todo este proyecto. A Laura Dubcovsky por estar siempre presente y tratarnos como familia. A mi familia por todo su apoyo.*

*En especial a mis dos amores... Tomy y Ale.*

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en ésta u otra institución.

*Publicaciones derivadas de la tesis:*

Nestor Kippes, Juan M. Debernardi, Hans Vasquez-Gross, Bala A. Akpinar, Budak Hikmet, Kenji Kato, Shiaoman Chao, Eduard Akhunov and Jorge Dubcovsky (2015) **Identification of the *VERNALIZATION 4* gene reveals the origin of spring growth habit in ancient wheats from South Asia.** Proc Natl Acad Sci USA 112 (39):E5401-E5410.

Nestor Kippes, Jie Zhu, Andrew Chen, Leonardo Vanzetti, Adam Lukaszewski, Hidetaka Nishida, Kenji Kato, Jan Dvorak, Jorge Dubcovsky (2013) **Fine mapping and epistatic interactions of the vernalization gene *VRN-D4* in hexaploid wheat** Mol Genet Genomics 289 (1):47-62.

## ÍNDICE GENERAL

ÍNDICE DE TABLAS.....	VII
ÍNDICE DE FIGURAS.....	VIII
ABREVIATURAS.....	XIV
RESUMEN.....	XV
ABSTRACT.....	XVI
 <b>CAPÍTULO 1 Introducción: el cultivo del trigo y la regulación genética de la floración</b>	<b>2</b>
1.1 INTRODUCCIÓN	3
1.1.1 EL CULTIVO DEL TRIGO EN LA INDIA	4
1.1.2 CONTEXTO MUNDIAL DE LA PRODUCCIÓN DE TRIGO	5
1.1.3 REGULACIÓN GÉNÉTICA DE LA FLORACIÓN EN CEREALES DE CLIMA TEMPLADO	6
1.1.4 <i>VRN-D4</i>	10
1.2 HIPÓTESIS	12
1.3 OBJETIVOS	13
1.3.1 Objetivo general	13
1.3.2 Objetivos específicos	13
 <b>CAPÍTULO 2 Mapeo de alta densidad y análisis de genes candidatos para <i>VRN-D4</i></b>	<b>14</b>
2.1 INTRODUCCIÓN	15
2.2 MATERIALES Y MÉTODOS	16
2.2.1 Material vegetal	16
2.2.2 Condiciones de crecimiento	17
2.2.3 Mapa genético y físico	17
2.2.4 Caracterización de <i>TaVIL-D1</i> y otros genes candidatos	18
2.2.5 Análisis de genética comparativa	19
2.2.6 Análisis de expresión mediante qRT-PCR	20
2.2.7 Interacciones epistáticas	21
2.3 RESULTADOS	22
2.3.1 Secuenciación del gen candidato <i>TaVIL-D1</i>	22
2.3.2 Mapa de alta densidad de la región candidata	23
2.3.3 Mapeo de <i>VRN-D4</i> usando líneas monotelosómicas y ditelosómica	26
2.3.4 Análisis de genes candidatos mediante genética comparativa	28
2.3.5 Epistasis entre <i>VRN-D4</i> y otros <i>loci</i> que controlan el requerimiento de vernalización	32
2.4 DISCUSIÓN	35
2.4.1 Mapeo de <i>VRN-D4</i>	35
2.4.2 Genética comparativa de la región candidata para <i>VRN-D4</i> en <i>B. distachyon</i> y arroz	36
2.4.3 Análisis de la región candidata de <i>VRN-D4</i>	38
2.4.4 Interacciones epistáticas de <i>VRN-D4</i>	39
2.5 CONCLUSIÓN	41

<b>CAPÍTULO 3 Identificación, validación y caracterización del gen candidato para <i>VRN-D4</i></b>	<b>42</b>
3.1 INTRODUCCIÓN	43
3.2 MATERIALES Y MÉTODOS	44
3.2.1 Estudio de expresión de genes mediante RT-QPCR	44
3.2.2 Búsqueda y análisis de mutaciones en la población de TILLING	47
3.2.3 Secuenciación de BACs y mapa físico	48
3.2.4 Marcadores para el sitio de inserción 5DS/5AL	50
3.2.5 RNA-EMSA	50
3.2.6 Análisis de estructura de la población	51
3.2.7 Diversidad genética	52
3.3 RESULTADOS	52
3.3.1 Caracterización de la línea portadora de <i>VRN-D4</i>	58
3.3.2 Validación de <i>VRN-D4</i> usando mutaciones inducidas	62
3.3.3 Mapa físico de la región de <i>VRN-D4</i>	69
3.3.4 Caracterización del alelo para hábito de crecimiento primaveral	73
3.3.5 Distribución geográfica de <i>VRN-D4</i>	83
3.4 DISCUSIÓN	83
3.4.1 Un genoma dinámico	84
3.4.2 El origen de <i>VRN-D4</i>	85
3.4.3 El efecto de <i>VRN-D4</i> en la diversidad genética de <i>T. aestivum</i> ssp. <i>sphaero</i>	87
3.4.4 Nuevos mecanismos de control del requerimiento de vernalización	91
3.5 CONCLUSIÓN	93
<b>CAPÍTULO 4 Discusión general y conclusiones</b>	<b>93</b>
4.1 DISCUSION GENERAL	94
4.1.1 Discusión de los objetivos alcanzados	96
4.1.2 Análisis de las hipótesis planteadas	98
4.2 CONCLUSIÓN GENERAL	101
<b>BIBLIOGRAFÍA</b>	<b>111</b>
<b>APENDICE</b>	<b>111</b>

## ÍNDICE DE TABLAS

<b>Tabla 2.1. Representación gráfica de los haplotipos de los recombinantes críticos.</b> A= TDF-J; B= CS <sub>5402</sub> , H= heterocigota, C= B o H, y D= A o H. Un cambio en el color de la celda indica un evento de recombinación. El tiempo a floración observado en la prueba de progenie de cada recombinante crítica esta mostrado al pie de la tabla, para cada clase genotípica, como la media $\pm$ error estándar de la media. (Arriba se indica el nombre del marcador utilizado para el test de progenie (TP)). S= significativo ( $P < 0.0001$ ), NS = non-significativo.....	p25
<b>Tabla 2.2. Selección de genes putativos anotados como factores de transcripción o genes relacionados con funciones en desarrollo.</b> .....	p31
<b>Tabla 2.3. Interacciones entre <i>VRN-D4</i> y otros alelos que regulan el requerimiento de vernalización.</b> Se muestra la probabilidad de los efectos principales y las interacciones en un ANOVA factorial 2 x 3, la proporción de la variación explicada por el modelo ( $R^2$ ), el grado de dominancia y el efecto relativo de una combinación alélica de dosis simple. ....	p33
<b>Tabla 3.1. Análisis de las mutaciones encontradas en <i>VRN-D4</i>.</b> .....	p60
<b>Tabla 3.2. Distribución geográfica de las entradas postuladas como portadoras de <i>VRN-D4</i> en estudios previos.</b> La presencia de los productos de los marcadores para los bordes de la inserción 5AL/5DS, y la presencia de la secuencia de RIP-3 <sup>VRN-D4</sup> es indicada como “+” y la ausencia como “-”. La detección del número de copias de <i>VRN-A1</i> (CNV) es descripta en la sección de materiales y métodos. La duplicación en tándem de <i>VRNA1</i> fue estudiada mediante la presencia/ausencia de un doble pico C/T en la posición 349 de la región codificante de <i>VRN-A1</i> como se ha descrito con anterioridad (Díaz et al. 2013). TDF, GABO, CS <sub>5402</sub> y TDC fueron incluidos como controles. Los nombres de las líneas corresponden a los publicados en estudios previos (Iwaki et al. 2000 y 2001). Los nombres comenzando con la letra ‘N’ corresponden a entradas del Instituto de Citología y Genética (Novosibirsk, Rusia)....	p75
<b>Tabla 3.3. Alelos primaverales presentes en las entradas de <i>T. aestivum</i> ssp. <i>sphaerococcum</i> utilizadas en este estudio.</b> Los marcadores para los alelos de <i>VRN-A1</i> , <i>VRN-B1</i> y <i>VRN-D1</i> fueron descriptos previamente (Yan et al. 2004 y Fu et al. 2005). <i>VRN-A1a</i> = duplicación del promotor y <i>VRN-B1a</i> = inserción de un retrotransposon en la secuencia promotora, ambas ligadas a hábito de crecimiento primaveral (Chu et al. 2011). Los nombres comenzando con ‘PI’ ‘CItr’ corresponden a entradas de GRIN y nombres comenzando con la letra ‘K’ al Instituto Vavilov (St. Petersburg, Rusia). La presencia de <i>VRN-D4</i> fue estudiada con los marcadores para los bordes de la inserción 5DS/5AL, la presencia del SNP A367C, y la presencia de la secuencia de RIP-3 <sup>VRN-D4</sup> .....	p77
<b>Tabla 3.4. Diversidad genética por genoma de las cinco subspecies de <i>T. aestivum</i>.</b> Se utilizó un total de 14,263 SNPs para los cuales se conoce su ubicación cromosómica seleccionados del conjunto inicial de marcadores de 16,371 SNPs identificados en el chip 90K iSelect Illumina (Wang et al. 2014) para estudiar la diversidad genética de cada genoma. ....	p79



## ÍNDICE DE FIGURAS

- Figura 1.1. *Triticum aestivum* spp. *sphaerococcum* cultivado en la provincia de Karnataka, India.** (A) Plantas cultivadas en condiciones de campo (B) Espigas (C) granos. (Mori et al. 2013).....p5
- Figura 1.2. Regulación genética del requerimiento de vernalización.** Antes de la vernalización los niveles de *VRN1* y *TaFT* son mantenidos bajos debido a altos niveles del represor *VRN2*, evitando así la ocurrencia de la floración durante el otoño. Durante la vernalización un incremento en los niveles de *VRN1* actúa como inhibidor de *VRN2*, liberando *TaFT* de su represión. Una inducción de *TaFT*, produce un aumento de los niveles de *VRN1* en hoja lo que a su vez retroalimenta el ciclo. *TaFT* viaja de las hojas a los ápices, donde induce un aumento de *VRN1*, como resultado se induce la transición del estadio vegetativo al reproductivo. Flechas verdes indican promoción, naranja indica represión, y gris representa una vía desfavorecida (Adaptado de Trevaskis et al. 2007 y Chen et al. 2014).....p10
- Figura 2.1. Uso de líneas ditelocentricas para el mapeo de *VRN-D4*.** (A) Bando-C del cromosoma 5D, cromosoma telocentrico 5DS y 5DL e iso-cromosoma 5DS. (B) Tiempo a floración de las plantas con diferentes dosajes del brazo corto del cromosoma 5D (5DS): 0 copias (mt5DL, dt5DL y nullisomicos 5D), una copia (mt5DS), dos copias (dt5DS y TDF-J), y cinco copias (dt-iso5DS + 5D). El número de plantas para cada clase está indicado en la base de cada barra y el error estándar de la media está representado en la barra de error. Los valores de significancia *P* son el resultado de dos t-test para las muestras con variancias desiguales con los datos transformados para recuperar la normalidad de los residuales (test Shapiro-Wilk no significativo). La clase dt-iso-5DS + 5D no fue utilizada en las comparaciones por estar representada por una única muestra. Nótese la aceleración de la floración cuando el brazo corto del cromosoma 5D está presente. ....p27
- Figura 2.2. Mapeo comparativo de la región candidata de *VRN-D4*.** (A) Mapa de alta densidad de *VRN-D4* en la población de trigo hexaploide TDF-J x CS<sub>5402</sub> (3.182 gametas). Las distancias están expresadas en cM. (B & C) Mapas de la región de *VRN-D4* en *B. distachyon* y arroz. Las distancias son expresadas en Mpb y están basadas en las posiciones actuales de los respectivos genomas. Las regiones centromérica no están dibujadas a escala. (B) Cromosoma 4 de *B. distachyon*. (C) Cromosoma 12 de arroz.....p28
- Figura 2.3. Comparación de los niveles de transcritos de *TaAGL31* en hojas de TDF (*Vrn-D4*) y TDC (*vrn-D4*) en plantas de tres y cinco semanas de edad.** Los niveles de transcritos están expresados como veces-*ACTINA* y fueron calculados de acuerdo a la fórmula  $2^{(CT_{ACTINA} - CT_{TaAGL31})}$ . Las barras de error representan el error estándar de la media calculado a partir de 6 réplicas biológicas. ....p32
- Figura 2.4. Interacciones entre *VRN-D4* y otros alelos que regulan el requerimiento de vernalización.** Para simplificar la visualización solo graficaron solo los valores de las plantas homocigotas para los diferentes alelos de vernalización segregando en cada población. A Interacciones entre *VRN-D4* y *VRN-A1*. B Interacciones entre *VRN-D4* y *VRN-B1*. C Interacciones entre *VRN-D4* y

*VRN-D1*. **D** Interacciones entre *VRN-D4* y *VRN-B3*. Todas las interacciones son significativas ( $P < 0.0001$ , basado en ANOVA  $2 \times 3$ ).....p34

**Figura 3.1. Caracterización de la línea TDF (*Vrn-D4*).** Niveles de transcritos de (A) *VRN-1* (todos los homólogos combinados), (B) *FT1* y (C) *VRN-2*. (D) Efecto en tiempo a floración de *VRN-D4*. El experimento fue terminado 114 días después de la siembra. (E) Niveles de transcripción de los tres diferentes homólogos de *VRN*. (F) Número haploide de copias de *VRN-A1*. Jagger, Munstertaler y Moldova fueron incluidos como controles para 1, 2 y 3 copias de *VRN-A1* respectivamente (Zhu et al. 2014). (A-E) Las medias fueron calculadas a partir de ocho replicas biológicas. (F) Media calculadas a partir de cuatro replicas biológicas. Las barras de error en todos los gráficos representan el error estándar de la media (SEM). TDC y CS<sub>5402</sub>: invernical, sin alelos de hábito primaveral. TDF: primaveral, *Vrn-D4*. TDD= primaveral, *Vrn-A1a*. BW (Bobwhite): primaveral, *Vrn-A1*, *Vrn-B1*, *Vrn-D1*. NS= no significativo, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , y \*\*\* =  $P < 0.001$ ).....p54

**Figura 3.2. Niveles de expresión en hoja de los tres homólogos de *VRN1*.** (A) Plantas de tres semanas de edad (B) Plantas de cinco semanas de edad. Los resultados a semana 1, que forman parte del mismo experimento, son presentados en la figura 3.1. Las barras representan la media calculada de ocho replicas biológicas y el error representa el error estándar de la media (SEM). El hábito de crecimiento y los alelos presentes se describen a continuación: TDC (invernical, *vrn-A1 vrn-B1 vrn-D1 vrn-D4*), TDF (primaveral, *vrn-A1 vrn-B1 vrn-D1 Vrn-D4*), TDD (primaveral, *Vrn-A1 vrn-B1 vrn-D1 vrn-D4*), y BW (Bobwhite, primaveral, *Vrn-A1 Vrn-B1 Vrn-D1 vrn-D4*). NS = no significativo, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , y \*\*\* =  $P < 0.001$ .....p55

**Figura 3.3. Detección del SNP A367C de *VRN-D4*.** (A) Cromatogramas de secuenciación mostrando el polimorfismo A367C en las líneas portadoras del alelo dominante *Vrn-D4* (TDF y Gabo) y su ausencia en las líneas portadoras del alelo recesivo de *vrn-D4* (CS<sub>5402</sub>, CS: Chinese Spring y TDC). (B) Cromatogramas mostrando el polimorfismo A367C en los recombinantes críticos de la población TDF x CS<sub>5402</sub> detallada en el Capítulo 2. En los estudios de progenie realizados para cada individuo recombinante se estableció si dicha progenie segrega para el hábito de crecimiento primaveral (segregante). Cuando todos los individuos de esa progenie florecieron sin diferencias significativas entre ellos y de forma similar al padre portador de *VRN-D4* se los considera no segregantes y de floración temprana (temprana). No se encontraron recombinantes donde todos los individuos de la progenie tuvieron una floración tardía.....p56

**Figure 3.4. Expresión de los parálogos de *VRN1* del brazo cromosómico 5AL (*VRN-A1*) y 5DS (temporalmente denominado *VRN-5DS*). RT-PCR con cebadores que amplifican ambos genes (*VRN-A1* y *VRN-5DS*).** Los dos transcritos pueden ser identificados mediante el polimorfismo A367C, el cual genera un sitio adicional de corte con la enzima de restricción *Bst*NI en la secuencia de *VRN-5DS*. Luego de la digestión con *Bst*NI, el fragmento de 251-pb detectado en *VRN-A1* (señalado con una cabeza de flecha) es cortado en dos fragmentos de 195 pb y 56 pb en *VRN-5DS* (flechas). En la línea TDF, donde *VRN-5DS* y *vrn-A1* están presentes, solamente los transcritos de *VRN-5DS* (variante “C” del SNP 367) son detectados hasta la tercer semana. A las cinco semanas, una banda tenue correspondiente a *VRN-A1*

también es detectada en TDF (variante “A” del SNP 367). TDC tiene un hábito de crecimiento invernal y por lo tanto no expresa *VRN1*. Se incluyeron tres réplicas biológicas para cada genotipo. *ACTINA* (Yan et al. 2003) fue usado como control de carga de la amplificación por PCR.....p57

**Figura 3.5. Análisis de las mutaciones inducidas generadas por TILLING en *VRN-D4*.**

(A) Estructura génica de *VRN-D4*, las flechas indican la localización del SNP A367C, y las mutaciones de TILLING (E158K y la mutación en el sitio de splicing) y la localización del SNP A367C. (B) Mediante un alineamiento múltiple de 268 secuencias que incluyeron los genes *API*, *AG*, *FUL*, *CAULIFLOWER*, *VRN1* y otras proteínas de la familia MAD-BOX se realizó la representación gráfica de la conservación de cada aminoácido. Se utilizó el sitio web de WebLogo (weblogo.berkeley.edu) que a partir del alineamiento genera la representación gráfica donde el tamaño de cada letra representa un aminoácido y el tamaño es proporcional al su frecuencia en esas posición. La región mostrada cubre las posiciones del aminoácido 115 al 175 ( en *VRN-D4*) dentro del dominio K-Box. Las flechas indican la posición donde se encuentran los cambios K123Q (natural de *VRN-D4*) y el producido mediante EMS E158K. (C) Caracterización de los mutantes de TILLING en tiempo a floración. Efecto de la mutación E158K en tiempo a floración. wt: homocigota para el alelo salvaje, het: heterocigota, E158K= homocigota para el alelo mutante. (D) Efecto de la mutación de splicing en tiempo a floración. Spl. = homocigota para la mutación de splicing. Las barras indican la media de al menos seis replicas biológicas. Las barras de error indican el error estándar de la media (SEM). NS = no significativo, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , y \*\*\* =  $P < 0.001$ ).....p61

**Figura 3.6. Representación gráfica de la región cromosómica de *VRN-D4* en el brazo corto del cromosoma 5D.**

(A) Ensamblado de la secuencia cromosómica CS-dt5DS (~680 kb) comparada con las bases de datos de secuencias de los brazos 5DS y 5AL (IWGSC), *T. urartu* y tres bases de datos de *Ae. tauschii*. Fueron anotados dos genes (*VRN-D4* y *Cyclin A2-1*= TRIUR3\_24995) y un gene hipotético TRIUR3\_08792. (B) Representación esquemática de la secuencia ensamblada: Amarillo= secuencia >99% idéntica a *Ae. tauschii*, Rojo= >99% idéntica al brazo 5AL; Negro = contigs de *Ae. tauschii* (<http://aegilops.wheat.ucdavis.edu/ATGSP/>), Rectángulos grises= BACs secuenciados, Amarillo= región >99% idéntica, Triángulos= inserción 5DS/5AL (C y D) Marcadores moleculares para la presencia de los sitios de inserción 5DS/5AL río arriba y río abajo de *VRN-D4* respectivamente. El producto de 1,440-pb indica la presencia de la 5DS/5AL río arriba de *VRN-D4* y la banda de 687-pb de tamaño corresponde a un control positivo de amplificación de la PCR (correspondiente al EST BJ315664). El producto de amplificación de 1,283-pb indica la presencia del sitio de inserción 5DS/5AL río abajo de *VRN-D4* y la banda de 534-pb corresponde a un control positivo de amplificación (EST BE606654).....p64

**Figura 3.7. Productos de amplificación de PCR de los genes flanqueantes a *VRN-A1* en ADN**

**purificado del brazo cromosómico 5DS.** Se utilizó ADN del cromosoma 5DS purificado mediante cartometría de flujo (Instituto de Botánica Experimental, Olomouc, Republica Checa) para estudiar si los genes cercanos a *VRN-A1* en el cromosoma 5AL (Yan et al. 2003) se encuentran también presentes en la inserción en el cromosoma 5DS. CS: Chinese Spring, 5DS: ADN del brazo 5DS, TDC: Triple Dirk C. Usando el mismo ADN para todos los marcadores solamente *VRN-A1* fue amplificado, indicando que los

genes *PHYC*, *CYS* y *AGLGI* no están presentes en el cromosoma 5DS. Las bandas múltiples en el gen *CYS* se deben a amplificaciones de las diferentes copias de este gen, ninguna presente en el ADN del brazo 5DS. Los genes *PHYC*, *CYS* y *AGLGI* tampoco fueron encontrados en la base de datos del cromosoma 5DS mediante BLAST, confirmado los resultados de PCR.....p65

**Figura 3.8. Amplificación de PCR de los sitios de inserción en los recombinantes críticos de la población TDF x CS<sub>5402</sub>.** (A) Marcador para la inserción 5AL/5DS río arriba de *VRN-D4* (1,440-bp band). (D) Marcador para la inserción 5AL/5DS río abajo de *VRN-D4* (1,283-bp band). Nótese que todas las líneas portadoras del SNP A367C muestran amplificación de las bandas de 1,440-pb y 1,283pb. H= segregante para el hábito de crecimiento (heterocigota *VRN-D4 vrn-D4*), temp.= toda la progenie tuvo floración temprana (homocigota *Vrn-D4*), tardía= toda la progenie tuvo floración tardía (homocigota *vrn-D4*), control= como control positivo de la amplificación de PCR se incluyó un par de cebadores adicionales para producir amplificación en todas las líneas.....p66

**Figura 3.9. Filogenia de *VRN-D4*. Arbol filogenético basado en la región genómica de ADN de las secuencias de *VRN-A1* / *VRN-D4*** (Neighbor-Joining utilizando 1000-pb río arriba del codón de inicio y 1000-pb río abajo del codón de stop de la transcripción) utilizando el programa MEGA6 (Tamura et al. 2013). Los números en los nodos de este árbol indican los valores de confianza basados en un 5000 submuestreos (% de árboles simulados con este nodo presente). Con excepción de *VRN-D4* todas las otras secuencias corresponden a *VRN-A1*. Los números indican los identificados correspondientes en GenBank, con excepción de la secuencia de *T. urartu*, en la cual se indica el número de contig en el ensamblado del proyecto genoma del mismo (plants.ensemble.org, GCA\_000347455.1).....p67

**Figura 3.10. Análisis de la secuencia de proteína de *VRN-D4*.** Alineamiento múltiple de la secuencia de proteína predicha para *VRN-D4* y las disponibles en bases de datos públicas *VRN1*. Resaltado en rojo se detalla el cambio de aminoácido característico de *VRN-D4* K123Q. El dominio K-Box se indica en color gris. La secuencia de *T. urartu* fue obtenida del sitio web plants.ensembl.org (scaffold60538). La secuencia completa de *VRN-A1* en la variedad Chinese Spring (GenBank KR422423) y *VRN-D4* (GenBank KR422424) fueron obtenidas en este estudio. ....p68

**Figura 3.11. Detalle de la región correspondiente a RIP-3 en el primer intrón de *VRN-1* y *VRN-D4*.** (A) El sitio de unión de GRP7 en *Arabidopsis* (Streitner et al. 2008) comparte 8 nucleótidos con la secuencia de RIP-3 donde se une TaGRP2. (B) La secuencia canónica de RIP-3 fue encontrada en *VRN-B1*, *VRN-D1* y *VRN-A<sup>m</sup>1* (*T. monococcum*), mientras que *VRN-A1*, *T. urartu* y todas las líneas de trigo tetraploide y hexaploide secuenciadas en la Tabla A.5 comparten el cambio de nucleótido T2783C. El haplotipo de RIP-3 con 3 SNPs fue encontrado en *VRN-D4* y en las secuencias de *VRN-A1* de las variedades Chinese Spring, Jagger y Claire, el cual fue asociado con un débil requerimiento de vernalización. Las flechas en RIP-3 indican la posición de los polimorfismos 2780, 2783 y 2784 (contando desde el ATG de *VRN-D4*). Los nombres de las secuencias incluyen los números de acceso de GenBank.....p70

**Figura 3.12. Unión de TaGRP2 a la región RIP-3 en el primer intrón de *VRN1/VRN-D4*.** (A) Expresión de GST-TaGRP2 en *E. coli*, ET: Extracto total, EP: elución de proteína. (B) Ensayo de movilidad electroforética de ARN (REMSA) mostrando la interacción entre TaGRP2 y la sonda de 27-nt

de ARN de RIP-3, usando la misma secuencia publicada previamente (Xiao et al 2014). Fueron analizadas dos concentraciones de TaGRP2 (1= 1 $\mu$ M y 2= 2 $\mu$ M) (C) REMSA mostrando las interacciones entre TaGRP2 con las tres diferentes sondas de ARN de 27-nt: RIP-3 presento una interacción fuerte, se observe una reducción en la interacción con la sonda RIP-3<sup>VRN-A1</sup> (SNP T2783C), y se detectó una interacción muy débil con la sonda RIP-3<sup>VRN-D4</sup> (SNPs G2780C, T2783C y C2784T, también presente en *VRN-A1* en Jagger, Claire y Chinese Spring). (D) Estructuras secundarias predichas para las sondas de 27-nt utilizadas mediante el programa RNA fold (Gruber et al. 2008). Las flechas indican la posición de los tres SNPs desde el codón de inicio de la traducción en *VRN-D4* (2780, 2783 y 2784).....p71

**Figura 3.13. Interacción de TaGRP2 con las sondas de ARN del sitio RIP-3.** (A) Sitio de unión de GRP2 en el primer intrón de *VRN1*. (B) Estructura secundaria predicha para la secuencia de ARN incluyendo 50 nt o 200 nt para cada lado del sitio RIP-3. (C) Estructura secundaria predicha de las tres sondas de ARN de 34-nt utilizadas (-L). RIP-3L: secuencia canónica, RIP-3<sup>VRN-A1-L</sup>: SNP T2783C, RIP-3<sup>VRN-D4-L</sup>: SNPs G2780C, T2783C y C2784T. Las flechas indican las posiciones de los SNPs estudiados en la estructura predicha y los colores representan la probabilidad del apareamiento de las bases. (D) Ensayo de movilidad electroforética del ARN (REMSA) mostrando la interacción de TaGRP2 con las tres sondas de ARN. Las interacciones se estudiaron utilizando dos concentraciones de proteína (1= 1  $\mu$ M y 2= 2  $\mu$ M). Una sonda RIP-3L sin marcar fue utilizada como competidor. (E) Las mismas sondas utilizadas en D fueron estudiadas a dos diferentes temperaturas de unión (calles 1-3= 4 °C y calles 4-7= 23 °C).....p73

**Figura 3.14. Sitio de colección geográfica de las entradas de *T. aestivum* ssp. *sphaerococcum* utilizadas en este estudio.** Los colores representan cada localización geográfica de acuerdo a las coordenadas disponibles en GRIN ([www.ars-grin.gov/npgs/searchgrin.html](http://www.ars-grin.gov/npgs/searchgrin.html)). Los puntos fueron posicionados de acuerdo a las coordenadas geográficas utilizando Google Maps ([maps.google.com](http://maps.google.com)). Todas las entradas poseen *VRN-D4* con excepción de las entradas CItr 8610 y CItr 10911, ambas colectadas en China del este (gris).....p80

**Figura 3.15. Diversidad genética por subespecie y cromosoma.** La diversidad genética fue calculada para cada locus (14,236 SNP) y cada subespecie usando la fórmula del Contenido de Información Polimórfica (PIC) descrita en la secciona de materiales y métodos. Estos valores de PIC fueron promediados para cada cromosoma, y para reducir los efectos de sesgos en el diseño de esta plataforma, se dividió este promedio por la diversidad genética promedio del genoma y la subespecie correspondiente. Nótese que el cromosoma 5D de *T. aestivum* ssp. *sphaerococcum* (*sphaero.*) es uno de los valores más bajos de diversidad genética, y este cromosoma muestra una diversidad genética más alta en las otras cuatro subespecies. ....p80

**Figura 3.16. Representación de la matriz de correlación de las 100 entradas de *T. aestivum* (5 subespecies) utilizadas en este estudio.** Este análisis fue basado en 16.371 SNPs. Rojo= alta similitud. Lo números a la derecha indican los identificadores mostrados en la Tabla A.6 y la Figura 3.17. (A) En la barra exterior se indica la presencia o ausencia de *VRN-D4* mediante color gris oscuro y claro

respectivamente. **(B)** En la barra de color se indican las entradas de: *T. aestivum* ssp. *aestivum* hábito de crecimiento primaveral en verde, y con hábito de crecimiento invernal en celeste, *T. aestivum* ssp. *compactum* en color naranja, *T. aestivum* ssp. *sphaerococcum* in en azul oscuro, *T. aestivum* ssp. *macha* en amarillo, y *T. aestivum* ssp. *spelta* con color rosado. Los asteriscos indican las dos entradas descriptas como *T. aestivum* ssp. *sphaerococcum* colectadas en China y que no poseen *VRN-D4* (Citr 8610 y Citr 10911). La clasificación taxonómica sigue se corresponde con la información disponible en GRIN ([www.ars-grin.gov/npgs](http://www.ars-grin.gov/npgs)).....p81

**Figura 3.17. Relaciones genéticas entre las subespecies de *T. aestivum* estudiadas.** **(A)** Gris oscuro indica la presencia y gris claro la ausencia de *VRN-D4* basado en los marcadores desarrollados en este estudio. **(B)** Estructura de la población basada en 16371 marcadores SNP. Cada barra representa un genotipo (100 líneas en total). **(C)** Clasificaciones taxonómicas de acuerdo a GRIN ([www.ars-grin.gov/npgs](http://www.ars-grin.gov/npgs)). Verde: *aestivum* (hábito de crecimiento primaveral), naranja: *compactum*, azul: *sphaerococcum*, amarillo: *macha*, rosa: *aestivum* (hábito de crecimiento invernal), cyan: *spelta*. Los asteriscos indican las dos entradas de clasificadas como ssp. *sphaerococcum* colectadas en China donde *VRN-D4* no está presente. Los identificadores de las entradas están descriptos en la Tabla A.6. **(D)** Nivel de diferenciación genética entre todas las combinaciones de pares de subespecies determinada mediante el cálculo de índice de diferenciación (*F<sub>st</sub>*). Todas las comparaciones son significativas con un  $P < 0.001$ . **(E)**. Valores de diversidad genética estandarizada para la región distal y proximal de los cromosoma del grupo 5. Nótese la alta reducción de la diversidad en la región proximal del cromosomas 5D en *T. aestivum* ssp. *sphaerococcum* relativa a las otras subespecies. ....p82

## ABREVIATURAS

ADNc: Ácido Desoxiribonucleíco Copia  
*AGL31: AGAMOUS LIKE 31*  
 ANOVA: Análisis de varianza  
 ARNi: Ácido Ribonucleíco de Interferencia  
 ARNm: Ácido Ribonucleíco Mensajero  
 BAC: Bacterial Artificial Chromosome  
 BLAST: Basic Local Alignment Search Tool  
 BLAT: Basic Local Alignment Tool  
 BW: Bob White  
*CO: CONSTANTS*  
 cM: centiMorgan  
 CNV: Copy Number Variation  
 CS: Chinese Spring  
 CS<sub>5402</sub>: Línea sintética de sustitución del cromosoma 5D en Chinese Spring  
 EST: Expressed Sequence Tag  
*FD: FLOWERING LOCUS D*  
*FLC: FLOWERING LOCUS C*  
*FRI: FRIGIDA*  
*FT: FLOWERING LOCUS T*  
*GRP2: GLYCINE-RICH RNA-BINDING PROTEIN 2*  
*GRP7: GLYCINE-RICH RNA-BINDING PROTEIN 7*  
 IWGSC: International Wheat Genome Sequencing Consortium  
 MTP: Minimum Tiling Path  
 PCR: Polimerase Chain Reaction  
*RHT: REDUCED HEIGHT GENE*  
 SNP: Single Nucleotide Polimorfism  
*SOC1: SUPPRESSOR OF OVEREXPRESSION OF CO 1*  
 TDB: Triple Dirk B (*Vrn-B1*, dominante primaveral)  
 TDC: Triple Dirk C (invernal)  
 TDD: Triple Dirk D (*Vrn-A1*, dominante primaveral)  
 TDE: Triple Dirk E (*Vrn-D1*, dominante primaveral)  
 TDF: Triple Dirk F (*Vrn-D4*, dominante primaveral)  
 TILLING: Targeting Induced Local Lesions IN Genomes  
*VRN1: VERNALIZATION 1*  
*VRN2: VERNALIZATION 2*  
*VRN3: VERNALIZATION 3*  
*VIL1: VERNALIZATION-INSENSITIVE 3 LIKE 1*

## Clonado posicional del gen *VRN-D4* en trigo hexaploide (*Triticum aestivum*)

### RESUMEN

Las variedades de trigo con hábito de crecimiento invernal requieren una prolongada exposición a bajas temperaturas para acelerar la floración. Este requerimiento se denomina ‘vernalización’ y la variación natural en los genes que controlan este proceso ha favorecido la adaptación del trigo a diferentes ambientes. Los principales *loci* que controlan el requerimiento de vernalización en trigo se denominan *VRN1*, *VRN2*, *VRN3* y *VRN-D4*. Los primeros tres genes han sido aislados y caracterizados pero *VRN-D4* no ha sido identificado aún a nivel molecular. El objetivo de esta tesis es la identificación de *VRN-D4* mediante clonado posicional, lo que incluyó la construcción de un mapa genético de alta densidad, un mapa físico, el estudio de los genes candidatos y su validación. Usando esta estrategia se encontró que *VRN-D4* está localizado en la región centromérica del brazo corto del cromosoma 5D. En esta región se identificó la inserción un fragmento cromosómico de ~290Kb del cromosoma 5A incluyendo una copia del gen de floración *VRN-A1*. Usando mutantes inducidos por EMS este estudio demuestra que esta segunda copia de *VRN-A1* es *VRN-D4*. *VRN-D4* fue encontrado en muy alta frecuencia en la sub-especie *T. aestivum* ssp. *sphaerococcum* predominante en el sur del continente Asiático. Esta sub-especie carece de otros genes para hábito de crecimiento primaveral, lo que sugiere que *VRN-D4* jugó un rol importante en la adaptación del trigo a esta región. Este estudio también identificó mutaciones en una región regulatoria de *VRN-D4* asociadas a su expresión temprana. Mutaciones similares identificadas en alelos de *VRN-A1* también estuvieron asociadas con un reducido requerimiento de vernalización. Los alelos de *VRN-A1* y *VRN-D4* identificados en este estudio representan una herramienta útil para regular la fecha de floración de trigo. Este estudio contribuye también al conocimiento básico de los mecanismos moleculares involucrados en la respuesta a la vernalización.

Palabras clave: trigo, floración, vernalización, *VRN1*, *VRN2*, *VRN3*, *VRN-D4*, *sphaerococcum*, *TaGRP2*, TILLING, genómica comparativa.



## Positional cloning of the *VRN-D4* gene in hexaploid wheat (*Triticum aestivum*)

### ABSTRACT

Wheat varieties with a winter growth habit require long exposures to low temperatures (vernalization) to accelerate flowering. Natural variation in the vernalization genes has favored wheat adaptation to different environments. The main wheat vernalization genes *VRN1*, *VRN2*, and *VRN3* have been cloned and characterized at molecular level but *VRN-D4* has remained unknown. The main objective of this thesis was to use the positional cloning approach to progress towards the identification of *VRN-D4*, aiming to produce a genetic and a physical map to identify candidate genes that can be study as potential candidates for *VRN-D4*. Using this strategy the *VRN-D4* was narrowed down to the centromeric region of the short arm of the chromosome 5D. In this region a ~290-kb DNA fragment from chromosome arm 5AL carrying an extra copy of the vernalization gene *VRN-A1* was detected. EMS induced mutations demonstrated that this gene is *VRN-D4*. *VRN-D4* was found in high frequency in the subspecies *Triticum aestivum* ssp. *sphaerococcum* from South Asia. This subspecies lack of any other allele for spring growth habit suggesting that *VRN-D4* played an important role in the adaptation of wheat to this area. This study also characterized mutations in regulatory regions of *VRN-D4* associated with its high expression that were also identified in *VRN-A1* alleles linked with reduced vernalization requirement. The alleles of *VRN-D4* and *VRN-A1* described in this study represent a new tool to regulate flowering time in wheat. This study also contributes to the basic understanding of the molecular mechanism involved in the vernalization response.

Key words: wheat, flowering, vernalization, *VRN1*, *VRN2*, *VRN3*, *VRN-D4*, *sphaerococcum*, *TaGRP2*, TILLING, comparative genomics.

## CAPÍTULO

## 1

**Introducción: el cultivo del trigo y la regulación  
genética de la floración.**

## 1.1 INTRODUCCIÓN

**E**l trigo pan (*Triticum aestivum* L.) es uno de los cereales más ampliamente cultivado en el mundo y provee alrededor de 20% de las calorías y proteínas consumidas mundialmente (FAO 2013). Esto lo hace un componente crítico de la alimentación humana y de la seguridad alimentaria mundial.

El cultivo del trigo ha sido fundamental para el hombre desde el origen de las primeras civilizaciones. La transición de los humanos de cazadores-recolectores a agricultores ocurrió en el cercano oriente (actuales Turquía, Siria, Israel, Jordania, Iraq e Irán) hace aproximadamente 12.000 años, y se estima que fue causada por los extensos cambios en las temperaturas y disponibilidad de agua que tuvieron lugar en ese período (Pleistoceno tardío), lo que motivó la identificación de nuevas y más estables fuentes de alimentos (Salamini et al. 2002). La domesticación del trigo pan tiene su origen en las costas sureste del mar Caspio hace alrededor de 8.000 años (Dubcovsky y Dvorak 2007). Estas formas domesticadas han sido el foco de una intensa selección, que resultaron en la modificación de muchas de sus características fisiológicas, como el color, sabor, hábito de crecimiento, reducción en la dispersión de semillas y rendimiento, por citar algunos (Diamond, J. 2002, Purugganan MD y Fuller DQ 2009).

El trigo pertenece al género *Triticum*, (tribu Triticeae, familia Poaceae). El género *Triticum* contiene varias especies que pueden ser clasificadas en cuanto a su número cromosómico: trigos diploides ( $2n=2x=14$ , genoma AA), tetraploides ( $2n=4x=28$ , genomas AABB y AAGG) y hexaploides ( $2n=4x=42$ , genoma AABBDD). Dentro de las especies más estudiadas se encuentran *T. monococcum* (diploide), *T. turgidum* ssp. *durum* (tetraploide, usado para pastas y cuscús) y *T. aestivum* (hexaploide usado

principalmente para pan). *T. aestivum* representa alrededor del 95% del trigo cultivado mundialmente, y el 5% restante corresponde a *T. turgidum* ssp. *durum* *T. monococcum* (conocido como eikorn wheat) el cual tiene una distribución marginal.

La forma hexaploide de trigo que se conoce actualmente es el fruto de dos procesos de hibridación espontánea que ocurrieron en las variedades diploides salvajes (Dubcovsky y Dvorak 2007).). En primer lugar, hace alrededor de 400.000 años, el trigo diploide salvaje ancestro del actual genoma A (*T. urartu*,  $2n = 2x = 14$ ) hibridizó con el ancestro del genoma B (una especie relacionada a *Aegilops speltoides*,  $2n = 2x = 14$ , genoma SS) para producir el trigo tetraploide salvaje conocido como ‘Wild Emmer’ (*T. dicoccoides*,  $2n = 4x = 28$ ) (Huang et al. 2002; Dvorak y Akhunov 2005). Este fue el primer trigo en ser cultivado y domesticado por el hombre y es el que dió origen al trigo pan mediante una segunda hibridación, esta vez con otro miembro del género *Aegilops*, que aportó el genoma D (*Ae. tauschii*,  $2n = 2x = 14$ , genoma DD) para producir las versiones actualmente utilizadas conocidas como *T. aestivum*.

*T. aestivum* incluye diferentes subespecies que fueron adoptadas en distintas partes del mundo y que son identificables por características morfológicas específicas: *T. aestivum* ssp. *compactum* (Europa oeste), *T. aestivum* ssp. *spelta* (Europa oeste), *T. aestivum* ssp. *macha* (Rusia) y *T. aestivum* ssp. *sphaerococcum* (India). Esta última subespecie es particularmente importante para este estudio debido a que el gen *VRN-D4* ha sido encontrado en algunas muestras de esta subespecie, sugiriendo que *VRN-D4* pudo haberse originado en esta subespecie en la antigua civilización del Valle del Indio (Goncharov y Shitova 1999, Goncharov 2003). Por lo tanto una revisión más detallada del desarrollo del trigo en India se incluye en la próxima sección.

### 1.1.1 EL CULTIVO DEL TRIGO EN LA INDIA

En la India el trigo se ha cultivado por más de 5.000 años. La subespecie *T. aestivum* subsp. *sphaerococcum* se cultivó principalmente en áreas del Norte de Pakistán e India con elevadas temperaturas y poca humedad. Esta subespecie se caracteriza por su baja estatura, floración temprana, resistencia a la roya amarilla y moderada resistencia a la sequía (Percival, 1921, Mori et al. 2013). *T. aestivum* subsp. *sphaerococcum* se encuentra distribuido principalmente en la región que comprende el Noreste de Pakistán

y Noroeste de la India (Punjab) por lo que se estima que este es su lugar de origen (Percival 1921, Ellerton 1939, Hutchinson 1976). *T. aestivum* subsp. *sphaerococcum* es distinguible de otras subespecies por su menor altura (semi-enano 60-70cm), tallos cortos y duros, barba corta, y granos semiesféricos (Figura 1.1). Esta variedad ha sido reemplazada por variedades modernas de *T. aestivum* ssp. *aestivum*, con mayor rendimiento (Hutchinson 1976), y su cultivo está actualmente limitado a pequeñas producciones para consumo local (Mori et al. 2013).

El interés en *T. aestivum* subsp. *sphaerococcum* ha crecido recientemente debido a que algunas de sus características, como la altura de planta y resistencia a enfermedades (Maccaferri et al. 2015), representan nuevas fuentes de variabilidad genética para el mejoramiento de los trigos modernos (Mori et al. 2013). Semillas clasificadas como *T. aestivum* subsp. *sphaerococcum* son abundantes en los registros arqueológicos de las antiguas ciudades de Harappa y Mohen-Daro de la civilización del Valle del Indio que floreció en esta región 2000-4000 años atrás. La distribución de estas muestras arqueológicas coincide con la distribución moderna de *T. aestivum* subsp. *sphaerococcum* (Weber 1999 y Weber et al. 2010).



**Figura 1.1. *Triticum aestivum* spp. *sphaerococcum* cultivado en la provincia de Karnataka, India.**  
(A) Plantas cultivadas en condiciones de campo (B) Espigas (C) granos. (Mori et al. 2013).

### 1.1.2 CONTEXTO MUNDIAL DE LA PRODUCCIÓN DE TRIGO

De acuerdo a las estimaciones más recientes se espera que la población mundial alcance aproximadamente los 9.000 millones de personas para el año 2050 (Godfray et al., 2010). Debido a que la expansión del área cultivable existente es insuficiente para

satisfacer el incremento esperado en la demanda, es necesario generar un incremento significativo en los rendimientos de trigo, y otros cereales (Parry et al., 2011; Reynolds et al., 2011).

La utilización de fertilizantes así como el mejoramiento genético y la mejora en el manejo de las enfermedades han producido un considerable incremento en el rendimiento en trigo durante el último siglo (Semenov et al., 2012). Sin embargo, los rendimientos de los cereales cultivados como el trigo son altamente sensibles a cambios climáticos y medioambientales. El calentamiento global, representa un desafío adicional para obtener los incrementos de rendimiento deseados (Semenov et al., 2012). Las estimaciones actuales predicen que para el año 2100 la temperatura media de la tierra puede incrementarse entre 1.1 y 6.4°C (Agencia de Protección del Medioambiente de los Estados Unidos, Global Climate Change Impacts in the United States Report 2009).

El trigo es particularmente sensible a temperaturas extremadamente frías o calientes especialmente durante el estadio reproductivo (Saini et al., 1983; Marcellos y Single, 1984; Alghabari et al., 2014; Vara Prasad y Djanaguiraman, 2014). Cambios en las temperaturas de cultivo pueden acelerar la transición a estadios reproductivos generando una exposición a temperaturas subóptimas para el desarrollo del grano, y por consiguiente a reducciones en el rendimiento (Fuller et al., 2007). En este contexto, un mejor entendimiento de las bases genéticas que regulan los estadios fisiológicos del desarrollo puede proveer herramientas adicionales para una optimización del ciclo reproductivo de trigo a nuevas condiciones ambientales.

### 1.1.3 REGULACIÓN GÉNÉTICA DE LA FLORACIÓN EN CEREALES DE CLIMA TEMPLADO

El cultivo de trigo está ampliamente distribuido en todo el mundo, actualmente se estima que es producido en más de 125 países (FAO, 2013). Su amplia distribución y su habilidad de adaptación a diferentes condiciones ambientales es atribuido en parte a la combinación de un genoma con una alta tasa de cambio (indels) combinado con un sistema poliploide que actúa como buffer de los cambios (Dubcovsky & Dvorak 2007) y en parte a la plasticidad de la red genética que regula su desarrollado reproductivo (Distelfeld et al. 2009a).

El tiempo a floración es crucial para el éxito reproductivo y, en el caso de los cereales cultivados, para maximizar la producción de semillas. En trigo el desarrollo reproductivo está controlado principalmente por el fotoperíodo y la vernalización. La variación natural en el requerimiento de fotoperíodo en trigo esta principalmente determinad por variantes alélicas en el locus *PPD1* (Welsh et al., 1973; Law et al., 1978; Scarth y Law 1983, Laurie et al., 1995). Las diferencias en el requerimiento de fotoperiodo tienen un impacto importante en la habilidad de trigo para adaptarse a diferentes ambientes. Por ejemplo, alelos que reducen la sensibilidad al fotoperíodo confiriendo floración temprana tanto en día largo (DL) como en día corto (DC) resultan beneficiosos para trigos invernales bajo condiciones agro-ambientales del continente Europeo (Worland 1996) y para el crecimiento en climas mediterráneos como el encontrado en California.

En combinación con las diferencias naturales en fotoperiodo, las diferencias en requerimiento de vernalización son las dos características principales en la adaptación del trigo a diferentes ambientes. Los trigos invernales requieren una prolongada exposición a bajas temperaturas para acelerar la transición de estado vegetativo a reproductivo (vernalización). Dicho requerimiento está ausente o se encuentra disminuido en trigos primaverales. La variación natural en el requerimiento de vernalización en trigo está determinada al menos por cuatro *loci*: *VRN1*, *VRN2*, *VRN3* y *VRN-D4* (Danyluk et al. 2003; Trevaskis et al. 2003; Yan et al. 2003, 2004a, 2006; Yoshida et al. 2010). Los genes responsables de estas diferencias han sido identificados a nivel molecular para los tres primeros *loci* (de los cuales se encuentran copias ortólogas en cada uno de los genomas que conforman el trigo: A, B y D). En los tres casos los alelos ancestrales son responsables por el tipo invernal mientras que los alelos determinantes de hábitos primaverales son el resultado de mutaciones en regiones reguladoras o codificantes que disminuyen el requerimiento de vernalización (Yan et al. 2004b, 2006; Fu et al. 2005; von Zitzewitz et al. 2005; Distelfeld et al. 2009b; Nitcher et al. 2013). El hábito de crecimiento primaveral determinado por *VRN1* y *VRN3* es dominante, en tanto que para *VRN2* es recesivo.

*VRN1* codifica para un factor de transcripción de tipo MADS-Box de alta similitud con el gen de identidad meristemática *API* de *Arabidopsis* (Danyluk et al. 2003; Trevaskis et al. 2003; Yan et al. 2003). La variación natural en la región promotora de *VRN1* así como en el primer intrón está asociada con una reducción o eliminación del requerimiento de vernalización y con una rápida inducción de los niveles de los

transcriptos de *VRN1* en ápices y hojas de plantas no vernalizadas (Yan et al. 2004b; Fu et al. 2005). Recientemente se ha identificado una región del primer intrón de *VRN1* como un sitio de unión a la proteína TaGRP2 (Xiao et al 2014). Esta región se encuentra dentro de la ‘región crítica’ definida en base a deleciones del primer intrón asociadas a floración temprana. *TaGRP2* es un homólogo de la proteína de unión a ARN de simple cadena *GRP7* la cual juega un rol importante en la regulación de la floración en *Arabidopsis* (Streitner et al. 2008). En ausencia de bajas temperaturas, TaGRP2 se une al pre-ARNm de *VRN1* e inhibe su expresión. Durante la vernalización TaGRP2 es gradualmente glicosilada (O-GlcNAc) permitiendo la unión de otra proteína denominada VER2. VER2 es una lectina jacalina que reduce la represión de la acumulación del ARNm de *VRN1* mediada por TaGRP2 (Xiao et al 2014).

A diferencia de *Arabidopsis*, la expresión de *VRN1* no está restringida al meristema apical, sino que también ocurre en las hojas. La principal función de *VRN1* en las hojas al principio de la primavera es mantener la represión de *VRN2*, un gene que opera como un represor de floración en día largo (Chen y Dubcovsky 2012).

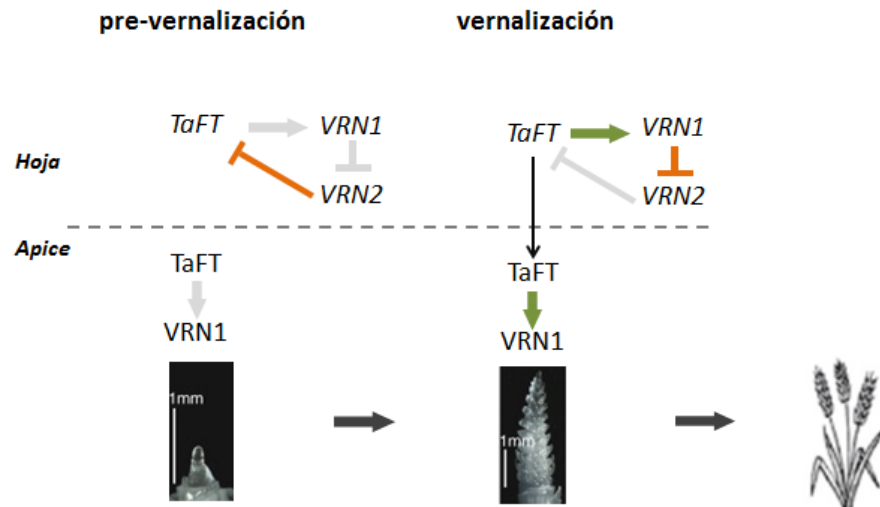
*VRN2* es un gene característico de los cereales templados ya que no posee un claro ortólogo en *Arabidopsis* (Yan et al. 2004a). Los homólogos más cercanos a *VRN2* descritos hasta la actualidad se encuentran en Sorgo y Arroz (Murphy RL et al. 2014, Xue et al. 2008), donde también actúan como represores durante día largo (*Ghd7*). La principal función de *VRN2* es actuar como un represor de *FT1* en el otoño para prevenir la floración antes del invierno, lo que podría dañar los meristemas florales.

*VRN3* (= *FT1*) en trigo es ortólogo a *FT* (*Flowering Locus T*) en *Arabidopsis* y actúa como un integrador de señales para la inducción de la floración. *TaFT* es inducido en día largo a actúa como un promotor de transcripción de *VRN1* (Yan et al., 2006; Hemming et al., 2008), Niveles altos de *FT1* en estadios tempranos están asociados con hábito de crecimiento primaveral (Yan et al. 2006).

La regulación de *VRN1*, *VRN2* y *FT1* esta interconectada formando un circuito de retroalimentación positiva. Antes de la vernalización este sistema se encuentra frenado por altos niveles del represor *VRN2*, este represor actúa sobre *FT1* previniendo la ocurrencia de la floración durante el otoño. Durante el período de vernalización se produce un incremento en los niveles de expresión de *VRN1*, el cual previene la inducción de *VRN2* cuando los días comienzan a alargarse en la primavera. La presencia de días largos en ausencia de *VRN2* permite la inducción de *FT1*. Un incremento en los niveles de *FT1* produce una inducción de los niveles de expresión de



*VRN1* retroalimentando el ciclo inductor de *FT1* y *VRN1* hasta que esos alcanzan los altos niveles de expresión característicos de la floración (Figura 1.2) (Chen et al. 2014).



**Figura 1.2. Regulación genética del requerimiento de vernalización.** Antes de la vernalización los niveles de *VRN1* y *TaFT* son mantenidos bajos debido a altos niveles del represor *VRN2*, evitando así la ocurrencia de la floración durante el otoño.

Durante la vernalización un incremento en los niveles de *VRN1* actúa como inhibidor de *VRN2*, liberando *TaFT* de su represión. Una inducción de *TaFT*, produce un aumento de los niveles de *VRN1* en hoja lo que a su vez retroalimenta el ciclo. *TaFT* viaja de las hojas a los ápices, donde induce un aumento de *VRN1*, como resultado se induce la transición del estadio vegetativo al reproductivo. Flechas verdes indican promoción, naranja indica represión, y gris representa una vía desfavorecida (Adaptado de Trevaskis et al. 2007 y Chen et al. 2014).

#### 1.1.4 *VRN-D4*

El clonado de los genes de vernalización *VRN1*, *VRN2* y *FT1* ha permitido un avance significativo de nuestro entendimiento del proceso de vernalización en los cereales tempranos. Sin embargo, un cuarto gen de vernalización, que es importante en Asia aún no ha sido identificado a nivel molecular.

Las primeras evidencias sobre la existencia de *VRN-D4* fueron encontradas por Knott en la década del '50 en Gabo (Knott et al. 1959), una variedad ampliamente cultivada en el período 1940 – 1960 en Australia. *VRN-D4* fue transferido a Gabo desde un cultivar Indio denominado Muzaffarnagar (O'Brien et al. 2001). Gabo fue utilizado más tarde

para producir una línea isogénica portadora de *VRN-D4* en el fondo genético Triple Dirk, denominada TDF (Pugsley 1972), a modo de estudiar las interacciones con otros alelos que regular el requerimiento de vernalización. Usando pruebas de alelismo, Stelmakh y Avsenin (1996) confirmaron que *VRN-D4* tienen carácter dominante y no es alélico a *VRN1* y, más tarde mediante estudios de segregación genética, fue detectado en *T. aestivum* subsp. *sphaerococcum* posicionándolo como la posible fuente de *VRN-D4* (Goncharov y Shitova 1999, Goncharov 2003). Los primeros mapas desarrollados para identificar *VRN-D4* lo ubicaron en el cromosoma 5D cercano al marcador microsatélite *gdm3* (Kato et al. 2003). Algunos estudios iniciales cuestionaron la existencia de *VRN-D4* y su localización cromosómica (Maystrenko 1980, Stelmakh 1987 y Goncharov 2003), sin embargo análisis posteriores pudieron confirmar su existencia (Goncharov y Shitova 1999, Yoshida et al. 2010). Mediante análisis genéticos el alelo para hábito de crecimiento primaveral de *VRN-D4* fue encontrado distribuido en muchas variedades de trigo primaverales de diferentes partes del mundo, y la frecuencia más alta fue identificada en el sur del continente Asiático (India y Pakistán) (Iwaki et al. 2000 y 2001). Estas evidencias sugieren que *VRN-D4* puede contribuir a la variación en floración en trigo y su identificación puede tener aplicaciones prácticas en el mejoramiento de trigo.

## 1.2 HIPÓTESIS

1. *VRN-D4* es un gen diferente a los genes de vernalización clonados anteriormente, y mutaciones en este gene afectaran la respuesta a vernalización en trigo y la duración del “*periodo a espigazón*”.
2. En el modelo de regulación de la etapa reproductiva de trigo, mediada por respuesta a vernalización, *VRN-D4* es reprimido, directa o indirectamente por *VRN2*, y su expresión promueve la expresión de *VRN1*.
3. La diversidad alélica de *VRN-D4* se encontrará asociada con variaciones en la secuencia de sus regiones regulatorias (promotor y/o intrón).
4. La relaciones de homeología existentes entre los genomas de trigo (A, B y D), permitirá identificar genes ortólogos a *VRN-D4* en los genomas A y B, y caracterizar la variabilidad existente en los mismos para ampliar las posibilidades de manipulación del desarrollo de la etapa reproductiva en programas de mejoramiento.

## 1.3 OBJETIVOS

### 1.3.1 Objetivo general

El objetivo a largo plazo de este proyecto es aumentar nuestro conocimiento sobre la regulación genética que determina el hábito de crecimiento en trigo, en particular, y en cereales de invierno, en forma general. Esto contribuirá a nuestro conocimiento básico sobre adaptación de los cultivos a diferentes ambientes y proveerá nuevas herramientas para manipular el ciclo de cultivo de trigo in diferentes condiciones medio ambientales.

### 1.3.2 Objetivos específicos

- 1 - Desarrollar un mapa genético de alta densidad para la región de *VRN-D4*.
- 2 - Construir un mapa físico de la región *VRN-D4* e identificar genes candidatos.
- 3 - Estudiar la expresión del gen candidato en relación a la de los genes *VRN1*, *VRN2* y *VRN3*.
- 4 - Estudiar la epistásis de *VRN-D4* con otros genes que regulan el requerimiento de vernalización como *VRN1*, *VRN2* y *VRN3*.
- 5 - Validar genes candidatos.

## CAPÍTULO

## 2

## Mapecto de alta densidad y análisis de genes candidatos para *VRN-D4*

---

Nestor Kippes, Jie Zhu, Andrew Chen, Leonardo Vanzetti, Adam Lukaszewski, Hidetaka Nishida, Kenji Kato, Jan Dvorak, Jorge Dubcovsky (2013) **Fine mapping and epistatic interactions of the vernalization gene *VRN-D4* in hexaploid wheat** Mol Genet Genomics 289 (1):47-62.

## 2.1 INTRODUCCIÓN

*VRN-D4* fue inicialmente identificado en el cultivar Australiano ‘Gabo’ (Knott 1959; Pugsley 1972). Este cultivar fue utilizado como donante para generar una línea isogénica en la serie desarrollada para el estudio de genes de vernalization en el fondo Triple Dirk y fue denominada ‘TDF’ (Pugsley 1972). Los primeros estudios de mapeo localizaron a *VRN-D4* en el cromosoma 5D (Kato et al. 1993), y más recientemente en un fragmento de 1,8 cM en la región centromérica del mismo cromosoma (Yoshida et al. 2010), sin poder precisarse el brazo cromosómico. Esta región de 1,8 cM incluye un gen denominado *TaVIL1* (Fu et al. 2007) el cual codifica para un homólogo del gen *VIL1* (VERNALIZATION INSENSITIVE3(VIN3)-LIKE 1) de *Arabidopsis*. *VIL1*, en conjunto con *VIN3* juega un papel importante en los cambios epigenéticos que ocurren durante la vernalización y también en la regulación de la floración por fotoperíodo (Sung et al. 2006). En trigo *TaVIL1* fue descrito como un gen cuyos niveles de transcripción se encuentra regulados por temperatura (Fu et al. 2007). En conjunto, estas evidencias de similitud con *VIL1* en *Arabidopsis*, localización y respuesta en los niveles de transcripción por vernalización hacen de *TaVIL-D1* un gen candidato para *VRN-D4*. En este capítulo se presentan los avances iniciales en el mapeo fino de la región de *VRN-D4* en el cromosoma 5D, el análisis de *TaVIL-D1* como gen candidato como así el de otros genes localizados en esta región.

## 2.2 MATERIALES Y MÉTODOS

### 2.2.1 Material vegetal

Se expandió la población de mapeo TDFxCS<sub>5402</sub>, de 159 líneas F<sub>2</sub> utilizadas por Yoshida et al. (2010) a 1500 plantas. TDF es una línea cuasi isogénica de trigo hexaploide, portadora del alelo primaveral para el gen *VRN-D4*, junto con los alelos recesivos para todas las copias del gen *VRN1* y *FT1*. *VRN1* posee tres homólogos que están ubicados en el brazo largo de los correspondientes cromosomas del grupo 5 (*VRN-A1*, *VRN-B1* y *VRN-D1*). CS<sub>5402</sub> es una línea de sustitución de trigo hexaploide, portadora del cromosoma 5D del trigo sintético 5402 en el fondo génico del cultivar Chinese Spring (CS). Este cromosoma 5D proviene de *Ae. tauschii* y porta el alelo

recesivo invernal para el locus *VRN-D4* y *VRN-D1*. *CS<sub>5402</sub>* tiene un hábito de crecimiento invernal debido al reemplazo del cromosoma 5D de CS, portador del alelo dominante en *VRN-D1*, que es el que otorga hábito primaveral al cultivar. Líneas nulli-tetrasómicas para el cromosoma 5D (nulisomía del cromosoma 5D compensada por tetrasomía del cromosoma 5A o del 5B, CSN5DT5A y CSN5DT5B, respectivamente) y líneas ditelosómicas para el brazo largo Dt5DL (líneas con ausencia del brazo corto del cromosoma 5D), disponibles en el cultivar Chinese Spring (Endo y Gill 1996; Linkiewicz et al. 2004; y Sears 1978) fueron utilizadas para determinar la ubicación cromosómica de los marcadores moleculares desarrollados para este estudio. La localización en el brazo cromosómico fue validada mediante la utilización de ADN de los brazos largo (5DS) y corto (5DL) del cromosoma 5D de Chinese Spring (Institute of Experimental Botany, Olomouc, Czech Republic), purificado mediante cartometría de flujo.

Para localizar *VRN-D4* en el correspondiente brazo cromosómico se produjeron líneas telocéntricas para el cromosoma 5D de la línea TDF mediante divisiones aberrantes de los cromosomas 5D univalentes generadas mediante el cruzamiento de líneas nulli-tetrasómicas CSN5DT5B con TDF. Las plantas F1 obtenidas fueron cultivadas y las 192 semillas obtenidas fueron analizadas mediante bandeo-C para detectar la presencia de productos de ruptura del cromosoma 5D (Lukaszewski y Gustafson 1983).

### 2.2.2 Condiciones de crecimiento

Todas las plantas de este estudio fueron cultivadas en invernáculo en condiciones no vernalizantes con una temperatura de entre 20–25 °C y en fotoperiodo largo (16h de luz). La longitud del día fue extendido mediante el uso de luz artificial para obtener un fotoperiodo deseado cuando no fue provisto por la luz natural. El tiempo a floración fue registrado como el tiempo desde la siembra hasta la emergencia de la espiga.

### 2.2.3 Mapa genético y físico

Los marcadores microsatélites utilizados para la búsqueda inicial de recombinantes (barc143, gdm3, wmc318, barc205, cfd67, cfd78 y cfd81) fueron obtenidos de la base de datos de marcadores del sitio web GrainGenes (<http://wheat.pw.usda.gov/GG2/>).

Las secuencias de los cebadores utilizados y las condiciones de los 19 marcadores basados en secuencias de ESTs están listados en la Tabla A.1. El marcador BG313707 ha sido descrito previamente (Yoshida et al. 2010). Los marcadores fueron asignados a los distintos brazos cromosómicos usando los materiales genéticos descritos en la sección de Material vegetal.

Las extracciones de ADN fueron llevadas a cabo usando el método de CTAB (Murray y Thompson 1980). Las reacciones de PCR fueron llevadas a cabo en un volumen final de 20 µl, conteniendo Tris-HCl 10 mM PH 8.3, KCl 50 mM, MgCl<sub>2</sub> 2 mM, dNTP 0.2 mM, 0.5 µM de cada cebadores, 50–100 ng de ADN , y 0.5 U Taq ADN polimerasa.

Las condiciones de amplificación utilizadas fueron las siguientes: 94°C por 5 min, seguido de 40 ciclos de 94°C por 30 s, “*annealing*” por 30 s, y 72°C por 1 min/kb, con una extensión final de 72°C por 7 min.

Las temperaturas de “*annealing*” para los marcadores fueron: 57°C para gdm68, cfd78 y barc143; 60°C para cfd67, gdm3 y wmc318; 65°C para bac205, y 60°C para TaVIL1. Los productos de PCR de los marcadores BE445181, BE499257, BE591275, BE606654, DQ512349(=TaAGL31), BE403761, BJ315664, CJ521028, y BE444702 fueron separados en geles de agarosa 1%. Todos los marcadores restantes y los marcadores microsatélites fueron separados en geles de poliacrilamida 6%. Todos los geles fueron visualizados con bromuro de etidio.

#### 2.2.4 Caracterización de *TaVIL-D1* y otros genes candidatos

La secuencia inicial de *TaVIL1* fue obtenida de *Triticum monococcum* (*TmVIL-A<sup>m</sup>1*) (Fu et al., 2007) y de la secuencia del BAC WCS1202110 (AB845602) perteneciente a la biblioteca de BACs de Chinese Spring (Allouis et al. 2003) usando cebadores para amplificar un fragmento inicial de *TaVIL-D1*. *TaVIL-D1* fue luego secuenciado en trigos diploides para obtener las secuencias correspondientes a los genomas A, B y D de trigo hexaploide a modo de utilizar los SNPs presentes únicamente en el genoma D para generar distintos juegos de cebadores que permitan amplificar toda la región del gen *TaVIL-D1*. Para esto se utilizaron *Ae. tauschii* (genoma D, AB845601), *Ae. speltoides* (genoma S, relacionado con el genoma B de trigo hexaploide, AB845599 y AB845600) y *Triticum urartu* (genoma A, AB845597 y AB845598). Utilizando las secuencias obtenidas se seleccionaron los SNPs genoma D específicos y se diseñaron cebadores que solo amplifiquen únicamente la copia del genoma D de trigo hexaploide. Las



condiciones de PCR y los cebadores utilizados para secuenciar *TaVIL-D1* están listadas en la Tabla A.2. La especificidad de los cebadores fue corroborada mediante la utilización de líneas nulli tetrasómicas para los cromosomas del grupo 5. Los productos de PCR genoma D específicos fueron secuenciados en las líneas TDF, CS<sub>5402</sub>, Triple Dirk C (TDC), Hayakomugi y Akakawaaka. La misma estrategia fue utilizada para secuenciar los otros genes candidatos. Una delección de 12pb detectada solo en CS<sub>5402</sub> fue utilizada para integrar *TaVIL-D1* en el mapa de alta densidad (cebador VIL1-D-F2 GTTGTTTCCTGTCCATACTAACGC y VIL1-D-R2 GGCTTTTTGTCTTGAAACATTTT).

### 2.2.5 Análisis de genética comparativa

El análisis de genética comparativa para delimitar la región candidata en *B. distachyon* que es colinear con la región de *VRN-D4* en trigo fue realizado mediante BLAST en el sitio web del genoma de *B. distachyon* (Brachy1.0, <http://www.brachypodium.org>) con los marcadores basados en ESTs que delimitan la región de *VRN-D4*. Los genes encontrados en *B. distachyon* fueron utilizados para buscar los genes ortólogos en arroz (NCBI, <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi> y Department of Energy's Joint Genome Institute and the Center for Integrative Genomics, <http://www.phytozome.net>). Si bien arroz no posee requerimiento de vernalización y está evolutivamente menos relacionado con trigo, su análisis permitió identificar genes colineares que sirvieron para diseñar nuevos marcadores para la región de mapeo. Para la búsqueda de potenciales candidatos para *VRN-D4*, el estudio fue enfocado en las comparaciones con *B. distachyon*. Los dominios proteicos de los genes encontrados en *B. distachyon* fueron anotados mediante la búsqueda de dominios conservados de NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) e InterProScan del EMBL-EBI (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>). De acuerdo con esta información se identificaron factores de transcripción y genes conocidos como involucrados en desarrollo. Estos genes fueron utilizados para buscar los homólogos más cercanos en trigo en la base de datos del NCBI, del proyecto genoma de trigo (International Wheat Genome Sequencing consortium, <http://urgi.versailles.inra.fr>) y de los transcriptomas más recientes (Krasileva et al. (2013), <http://maswheat.ucdavis.edu/Transcriptome/>). Los genes de trigo detectados fueron usados para buscar nuevamente en *B. distachyon*

usando la secuencia de proteína predicha mediante BLASTp, y de esta forma confirmar que el homólogo correcto fue encontrado.

#### 2.2.6 Análisis de expresión mediante qRT-PCR

Tejido de la última hoja completamente extendida fue colectado en nitrógeno líquido y el ARN fue extraído de acuerdo a las especificaciones del sistema Spectrum Plant Total RNA Kit (Sigma-Aldrich). La integridad del ARN en las muestras fue comprobada mediante geles de agarosa 1%. Dicho ARN se cuantificó usando un espectrofotómetro en la relación de absorbancia 260/280nm (NanoDrop ND-1000 Spectrophotometer). La expresión del gen candidato *TaAGL31* fue analizada con los cebadores F-CCAGATGCTGCAAAGCACTA y R-GAGCTCATTCTTCCTGGCC, estos cebadores están diseñados sobre regiones conservadas de los homólogos del genoma A, B y D. Para evitar amplificación de ADN genómico el cebador reverso fue diseñado sobre una unión de exones. La especificidad fue comprobada mediante curva de desnaturalización de las curvas estándar utilizadas para calcular la eficiencia de estos cebadores usando el sistema SYBR Green®. La reacción de PCR cuantitativa fue llevada a cabo en un equipo 7500 Fast Real-Time PCR system (Applied Biosystems). Como control endógeno de expresión génica se utilizó ACTINA (Uauy et al. 2006). Los niveles de expresión fueron calculados mediante la fórmula  $2^{(CT_{actina}-CT_{blanco})}$  y expresados en relación a los niveles de actina.

#### 2.2.7 Interacciones epistáticas

Se desarrollaron cuatro poblaciones segregantes para estudiar las interacciones epistáticas entre *VRN-D4* y otros genes que controlan el requerimiento de vernalización. La línea TDF fue cruzada con las líneas isogénicas portadoras de los alelos primaverales de *VRN-A1*, *VRN-B1* y *VRN-D1* (TDD, TDB y TDE respectivamente). Estas líneas fueron producidas para el estudio de los diferentes *loci* que controlan el requerimiento de vernalización (Pugsley AT 1971 y 1972). Una población adicional fue generada mediante el cruzamiento de TDF con la línea portadora del alelo primaveral de *TaFT1* CS(Hope7B) (Yan et al. 2006) para estudiar la interacción entre *VRN-D4* y *VRN-B3* (*TaFT1*). El cromosoma 7B proveniente de la variedad ‘Hope’ posee un alelo de *VRN-*

*B3* sobre expresado por la inserción de un elemento repetitivo en su región promotora (Yan et al. 2006). Aproximadamente 100 plantas para cada población F2 fueron cultivadas en invernáculo en condiciones no vernalizantes y día largo (24°C de temperatura y 16h de luz). Los genotipos de *VRN-A1*, *VRN-B1*, *VRN-D1* y *VRN-B3* fueron determinados mediante PCR mediante la utilización de marcadores diagnóstico previamente descritos (Fu et al. 2005; Yan et al. 2006). Los genotipos de *VRN-D4* fueron inferidos mediante el marcador microsatélite *barc205* localizado a 1.2cM distal a *VRN-D4* (Yoshida et al. 2010).

Se registró el número de días desde la siembra hasta la emergencia de la primer espiga para cada población F2. Los datos fueron evaluados usando ANOVA factoriales 2x3 considerando *VRN-D4* y cada uno de los otros *loci* como factores con tres niveles: homocigota *Vrn-D4*, heterocigota y homocigota *vrn-D4*. En el caso de la población *VRN-B3*x*VRN-D4* se consideró la segregación de otro locus dominante primaveral (*Vrn-D1*) proveniente de la línea CS(Hope7B), de esta manera solamente 33 plantas que poseían el alelo invernal de *VRN-D1* fueron consideradas en el análisis de las interacciones epistáticas.

Los análisis estadísticos fueron llevados a cabo con el programa SAS versión 9.3. El grado de dominancia fue calculado usando la siguiente formula:  $D = (2X_2 - X_1 - X_3)/(X_1 - X_3)$  (Falconer 1960) donde  $X_1$ ,  $X_2$  y  $X_3$  son los datos de tiempo a floración (HT) de las plantas homocigotas para el alelo invernal, heterocigotas y homocigotas para el alelo primaveral respectivamente. También se comparó el efecto relativo de *VRN-D4* con respecto a los otros genes de vernalización (*VrnX*) en tiempo a floración (HT) cuando cada uno de estos estaba presente en estado de heterocigosis en un fondo genético invernal según la siguiente formula:

Efecto relativo de dosis simple =

$HT(VrnD4/vrnD4\ vrnX/vrnX) - HT(VrnX/vrnX\ vrnD4/vrnD4)$ .

## 2.3 RESULTADOS

### 2.3.1 Secuenciación del gen candidato *TaVIL-D1*

La región génica que contiene a *TaVIL-D1* fue secuenciada en la línea portadora de *VRN-D4* TDF (No. Identificador en GenBank AB846583), en cinco líneas que no

poseen *VRN-D4*: Hayakomugi (AB846584), CS<sub>5402</sub> (AB846585), TDC (AB846586), Akakawaka (AB846587) y en *Ae. tauschii* (AB845601). La región secuenciada cubre desde 1 Kb río arriba del codón de inicio de la transcripción hasta 1.2 Kb río abajo del codón de terminación, proveyendo un total de 5.4 Kb. Todas las líneas donde *TaVIL-D1* fue secuenciado, con excepción de TDF, poseen un hábito de crecimiento invernal.

La línea denominada CS<sub>5402</sub>, es una línea de sustitución donde el cromosoma 5D de Chinese Spring fue reemplazado por el cromosoma 5 de *Ae. Tauschii* (progenitor del genoma D en trigo hexaploide) y fue la línea que mostró más polimorfismos. CS<sub>5402</sub> posee 18 cambios de nucleótidos (SNPs) en comparación con TDF. Dentro de estos cambios solo seis corresponden a cambios en regiones codificantes y cinco de estos producen cambios de aminoácidos. Adicionalmente se encontró una delección de 12 pb en el exón 2 que resulta en la pérdida de cuatro aminoácidos localizados fuera de los dominios conservados de la proteína (PHD, FNIII o VID). No se encontró ningún polimorfismo entre TDF y las otras líneas con hábito de crecimiento invernal (Hayakomugi, TDC, o Akakawaaka). La evidencia provista por la secuenciación de *TaVIL-D1* sugirió que este no es un buen candidato para *VRN-D4*, y para proveer una evidencia más concluyente se produjo una población de mapeo entre TDF y CS<sub>5402</sub> para avanzar en la delimitación de la región candidata mediante la búsqueda de eventos de recombinación entre *VRN-D4* y *TaVIL-D1*.

### 2.3.2 Mapa de alta densidad de la región candidata

Usando una población de mapeo de 1.591 plantas F2 totales (3,182 gametas) *TaVIL-D1* pudo ser mapeado en la región centromérica del brazo corto del cromosoma 5D (5DS). Utilizando las secuencias disponibles del proyecto de genoma de trigo hexaploide (<http://wheat-urgi.versailles.inra.fr/Seq-Repository/BLAST>) también encontramos los homeólogos de *TaVIL-D1* en los cromosomas 5AS y 5BS, y en la misma ubicación en cebada ([plants.ensembl.org](http://plants.ensembl.org)), sugiriendo conservación en la tribu de los *Triticeae*. Mediante un análisis de genómica comparativa también localizamos ortólogos en el cromosoma 4 de *Brachypodium distachyon* (organismo modelo en cebada y trigo) (Brad4g05950, 4.91 Mb) y en el cromosoma 12 de arroz (Os12g34850, 21.21 Mb).

La población de mapeo utilizada mostró una recombinación limitada pero a pesar de este fenómeno, se encontraron dos eventos de recombinación entre *VRN-D4* y *TaVIL-D1* (Tabla 2.1, plantas 189 y 768). Ensayos de progenie de estas dos recombinantes críticas demostraron que segregan para el carácter tiempo de floración cuando están fijas para *TaVIL-D1* (Tabla 2.1).

Estos dos eventos de recombinación independientes, en conjunto con la ausencia de polimorfismo a nivel de secuencia entre las líneas invernales y primaverales analizadas indican que *TaVIL-D1* es un muy poco probable candidato para *VRN-D4*.

Para caracterizar más detalladamente esta región se desarrollaron marcadores basados en secuencias expresadas cortas (ESTs) usando genes en las regiones colineares de arroz y *B. distachyon*. Usando estos nuevos marcadores se identificaron los eventos de recombinación detallados en la Tabla 2.1, donde *VRN-D4* fue mapeado en un intervalo de 0.09 cM delimitado por los marcadores BE404490 y BJ315664 en la región centromérica del cromosoma 5D.

Los nueve marcadores completamente ligados al centrómero no pudieron ser ordenados mediante los eventos de recombinación encontrados, pero fueron divididos en dos grupos en base a su localización en el correspondiente brazo cromosómico, determinada mediante el uso de líneas telocéntricas. *VRN-D4* co-segregó con tres marcadores mapeados en el brazo corto y seis marcadores correspondientes al brazo largo del cromosoma 5D, y por lo tanto no puede ser asignado a un brazo cromosómico. Para obtener esta información se produjeron líneas telocéntricas del cromosoma 5D de la línea portadora de *VRN-D4* (TDF).

Marcador	Recombinantes Críticos					
	no.189	no.366	no.768	no.1423	no.1279	no.1572
<i>BQ167501</i>	A	H	B	B	A	A
<i>cf81</i>	A	H	-	B	-	H
<i>cf78</i>	A	-	B	-	A	-
<i>BE499257</i>	A	H	B	B	A	H
<i>BE445181</i>	A	H	B	B		H
<i>TaVil-D1</i>	A	H	B	H	A	H
<i>BG313707</i>	D	D	B	-		-
<i>BE404490</i>	A	H	B	H	A	H
<i>BE606654</i>	H	H	H	H	A	H
<i>BG263382</i>	H	H	H	H	A	H
<i>TaAGL31</i>	H	H	H	H	A	H
<b>Vrn-D4</b> (floración)	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>A</b>	<b>H</b>
<b>Centrómero</b>						
<i>BE403761</i>	H	H	H	H	A	H
<i>CJ717651</i>	H	H	H	H	A	H
<i>CJ930261</i>	H	H	H	H	A	H
<i>CJ680715</i>	-	H	H	H	A	H
<i>cf67</i>	H	H	H	H	A	H
<i>CJ522104</i>	H	H	H	H	A	H
<i>BJ315664</i>	H	H	H	H	H	H
<i>CJ521028</i>	H	B	H	H	H	B
<i>TaFTL-1</i>	H	B	H	H	H	B
<i>BE444702</i>	H	B	H	H	H	B
<i>BE444353</i>	H	B	H	H	H	B
<i>BE405060</i>	H	B	H	H	H	B
<i>barc205</i>	H	B	D	D	D	B
<i>wmc318</i>	H	B	H	H	H	B
<i>gdm3</i>	H	B	C	C	C	C
<i>barc143</i>	H	B	H	H	H	B
<b>genotipo PP</b>	<b><i>cf67</i></b>	<b><i>cf67</i></b>	<b><i>cf67</i></b>	<b><i>barc205</i></b>	<b><i>BJ315664</i></b>	<b><i>BE606654</i></b>
A	85.7±2.51	83.4±3.31	66.3±2.60	73.8±2.04	51.6±0.61	63.4±3.37
B	139.0±2.00	134.7±2.53	132.0±2.30	103.4±1.48	58.1±2.84	118.5±3.90
H	116.2±5.46	107.4±4.50	106.5±2.33	97.3±1.94	57.1±1.76	102.0±4.65
P ANOVA	S	S	S	S	NS	S

**Tabla 2.1. Representación gráfica de los haplotipos de los recombinantes críticos.** A= TDF-J; B= CS<sub>5402</sub>, H= heterocigota, C= B o H, y D= A o H. Un cambio en el color de la celda indica un evento de recombinación. El tiempo a floración observado en la prueba de progenie de cada recombinante crítica esta mostrado al pie de la tabla, para cada clase genotípica, como la media ± error estándar de la media. (Arriba se indica el nombre del marcador utilizado para las pruebas de progenie (PP)). S= significativo ( $P < 0.0001$ ), NS = non-significativo.

### 2.3.3 Mapeo de *VRN-D4* usando líneas monotelosómicas y ditelisómica

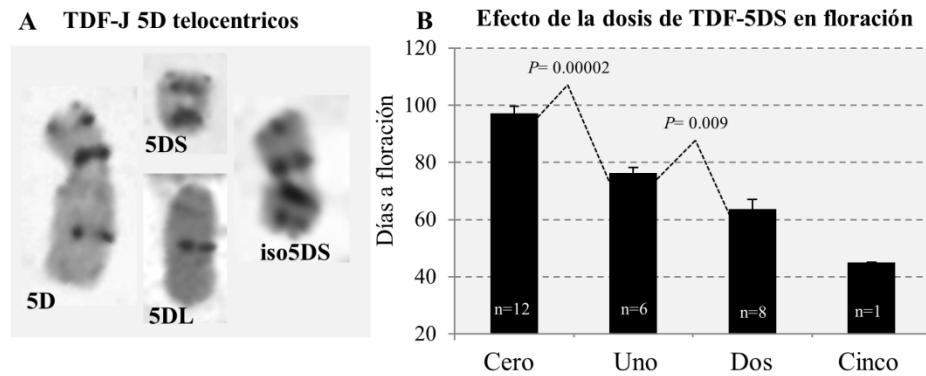
Para generar plantas monotelosómicas se produjo un cruzamiento entre la línea portadora de *VRN-D4* y una línea aneuploide careciente del cromosoma 5D . Se analizaron 192 plantas de esa progenie, en búsqueda de productos de divisiones fallidas del cromosoma 5D mediante el método de bandeo-C.

Se identificaron dos individuos monotelosómicos 5DL (brazo largo), uno para el brazo corto (mt5DS) y tres con un cromosoma 5D normal más un isocromosoma 5DS (fusión centromérica de dos brazos cortos) (Figura 2.1.A). Estas plantas fueron cultivadas y sus progenies analizadas para la presencia de cromosomas telocéntricos y tiempo a floración. Debido a que *VRN-D4* es de carácter semi-dominante para tiempo a floración (Yoshida et al. 2010) las plantas que no presenten aquel brazo cromosómico donde *VRN-D4* está ubicado deberían mostrar un retraso en el carácter de tiempo a floración.

Doce plantas donde el brazo corto del cromosoma 5D está ausente (tres mt5DL, cuatro dt5DL y cinco nulisómicas 5D) florecieron en  $97.1 \pm 2.6$  días después de siembra, lo que representa un retraso de 20.8 días con respecto a las líneas que poseen una copia del cromosoma 5DS cromosoma ( $76.3 \pm 1.9$ ,  $P = 0.00002$ ) (Figura 2.1.B). La comparación entre plantas que poseen una dosis del brazo corto (mt5DS) y aquellas que poseen dos (dt5DS y controles TDF) mostro que estas últimas produjeron una aceleración de 12.7 días en el tiempo a floración ( $63.6 \pm 3.5$ ,  $P = 0.009$ ) con respecto a las plantas que solo poseen una dosis del 5DS.

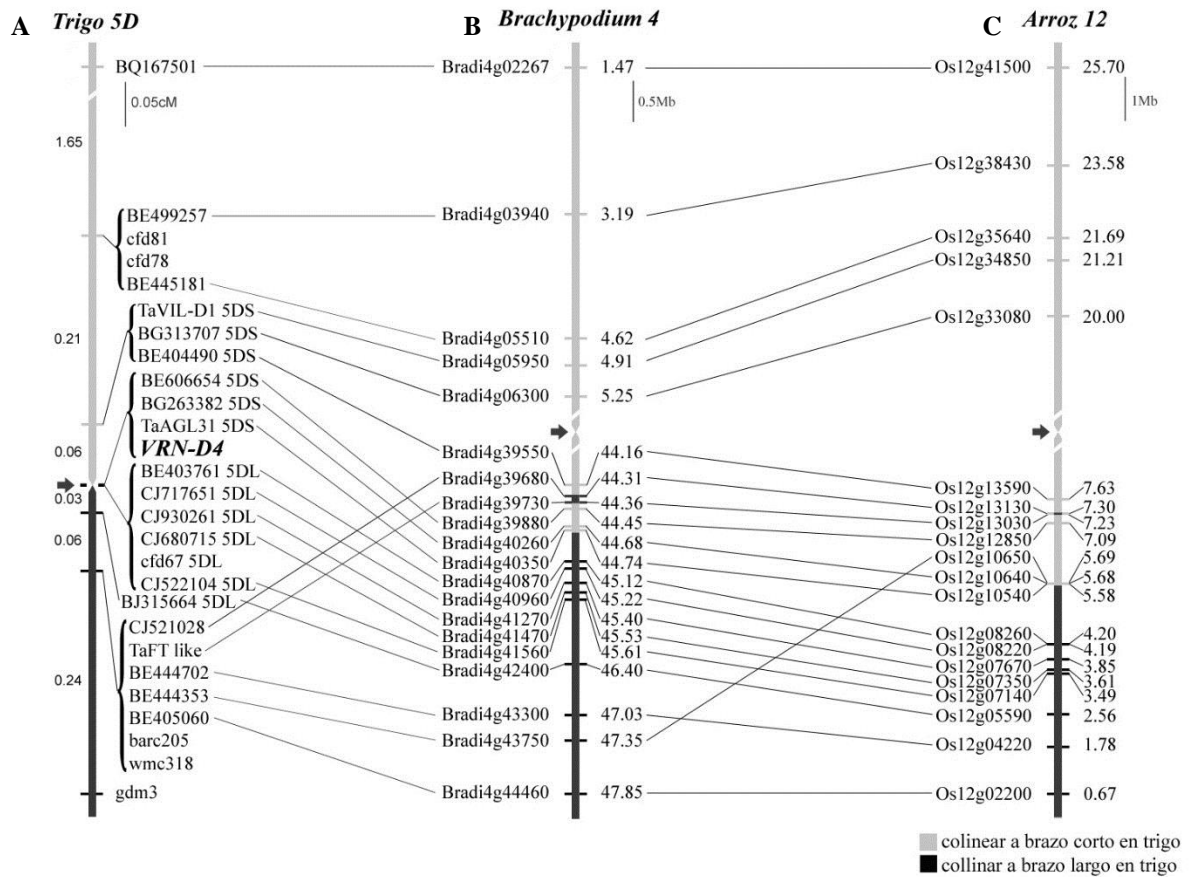
No se realizaron comparaciones estadísticas con una única planta que posee cinco dosis del brazo corto del cromosoma 5D, pero es interesante destacar que esta fue la planta que tuvo un tiempo a floración más corto (45 días) de todas las estudiadas.

En resumen, estos resultados demuestran que *VRN-D4* está localizado en el brazo corto del cromosoma 5D.



**Figura 2.1.** Uso de líneas ditelocentricas para el mapeo de *VRN-D4*. (A) Bando-C del cromosoma 5D, cromosoma telocentrico 5DS y 5DL e iso-cromosoma 5DS. (B) Tiempo a floración de las plantas con diferentes dosajes del brazo corto del cromosoma 5D (5DS): 0 copias (mt5DL, dt5DL y nullisomicos 5D), una copia (mt5DS), dos copias (dt5DS y TDF-J), y cinco copias (dt-iso5DS + 5D). El número de plantas para cada clase está indicado en la base de cada barra y el error estándar de la media está representado en la barra de error. Los valores de significancia  $P$  son el resultado de dos t-test para las muestras con variancias desiguales con los datos transformados para recuperar la normalidad de los residuales (test Shapiro-Wilk no significativo). La clase dt-iso-5DS + 5D no fue utilizada en las comparaciones por estar representada por una única muestra. Nótese la aceleración de la floración cuando el brazo corto del cromosoma 5D está presente.





**Figura 2.2. Mapeo comparativo de la región candidata de *VRN-D4*.** (A) Mapa de alta densidad de *VRN-D4* en la población de trigo hexaploide TDF-J x CS<sub>5402</sub> (3.182 gametas). Las distancias están expresadas en cM. (B & C) Mapas de la región de *VRN-D4* en *B. distachyon* y arroz. Las distancias son expresadas en Mpb y están basadas en las posiciones actuales de los respectivos genomas. Las regiones centromérica no están dibujadas a escala. (B) Cromosoma 4 de *B. distachyon*. (C) Cromosoma 12 de arroz.

### 2.3.4 Análisis de genes candidatos mediante genética comparativa

Los marcadores que delimitan la región candidata de *VRN-D4* fueron utilizados para delimitar regiones en *B. distachyon* (<http://www.brachypodium.org>) y arroz (<http://www.phytozome.net>). Los siete grupos de marcadores que incluyen veinte marcadores basados en ESTs de trigo mostraron buena colinearidad con excepción de dos marcadores (CJ521028 y *TaFT-like*). Todos estos genes (excepto Bradi4g43750) son colineales entre *B. distachyon* y arroz. A pesar de la buena colinearidad observada, el centrómero del cromosoma 5D en trigo fue mapeado en un región no colinear. El centrómero de trigo fue mapeado entre los marcadores *TaAGL31* y BE403761, 0.06 cM proximal a la posición del centrómero predicho en *B. distachyon* y arroz (Figura 2.2.A).

Los marcadores flanqueantes del centrómero en trigo son colineares con una región del cromosoma 4 de *B. distachyon* (44.74–45.12 Mb) que está localizada a 22 Mb del centrómero activo (22.5–22.9 Mb) (Qi et al. 2010) (Figura 2.2.B). Estos mismos marcadores de trigo son colineares con una región del cromosoma 12 de arroz (4.21–5.58 Mb) que está ubicada a 6.4 Mb del centrómero activo de este cromosoma en arroz (12 Mb) (Qi et al. 2010) (Figura 2.2.C).

Utilizando esta región delimitada en *B. distachyon* analizamos los genes predichos en esta región. Para esto se utilizaron los marcadores BE404490 (en el brazo corto) y BE403761 (en el brazo largo). Esta región contiene la región candidata en el brazo corto, el centrómero y posiblemente alguna porción del brazo largo ya que se utilizó un marcador del brazo largo para asegurar la cobertura total de la región centromérica.

En total la región candidata en *B. distachyon* comprende alrededor de 1 Mb (44.16–45.13 Mb), dicha región incluye 127 genes putativos (<http://www.brachypodium.org>). Dentro de estos genes, diez han sido anotados como factores de transcripción o con funciones relacionadas a desarrollo (Tabla 2.2) y son discutidos en detalle a continuación.

Tres de los genes seleccionados en *B. distachyon* (*Bradi4g39630*, *Bradi4g40057*, y *Bradi4g40517*) mostraron alta homología con genes localizados en el brazo largo del cromosoma 5D en trigo (según los resultados de BLAST y su anotación utilizando el borrador del genoma de trigo en [wheatgenome.org](http://wheatgenome.org)), por lo tanto fueron descartados como candidatos para *VRN-D4*. Este resultado no es sorprendente ya que se utilizó un marcador del brazo largo para asegurar la inclusión de toda la región centromérica del cromosoma 5D.

Dos de los genes candidatos en *B. distachyon* (*Bradi4g40250* y *Bradi4g40310*), no mostraron buena homología con los genes de trigo, y los encontrados están localizados en los cromosomas 6D y 7D respectivamente (Tabla 2.2). Por lo tanto, fueron descartados como candidatos para *VRN-D4*. El gen *Bradi4g40540* mostro homología con una secuencia en el cromosoma 5BS (IWGSC\_chr5BS\_ab\_k71\_2249326) pero no se encontraron homólogos en los cromosomas 5AS y 5DS (<http://wheaturgi.versailles.inra.fr/Seq-Repository>). La anotación de esta región genómica, reveló la presencia de un codón de STOP prematuro en el primer intrón, por lo que es altamente probable que se trate de un pseudogen y por lo tanto, no representa un buen candidato para *VRN-D4*.

Tres genes de *B. distachyon* (Bradi4g40230, Bradi4g40270, y Bradi4g40357) fueron descartados porque no presentaron homólogos en ninguno de los borradores de los distintos cromosomas del proyecto de genoma de trigo. BLAST reverso, utilizando las secuencias de trigo con baja homología detectadas para búsquedas de genes en *B. distachyon* detectaron genes distintos de los usados originalmente.

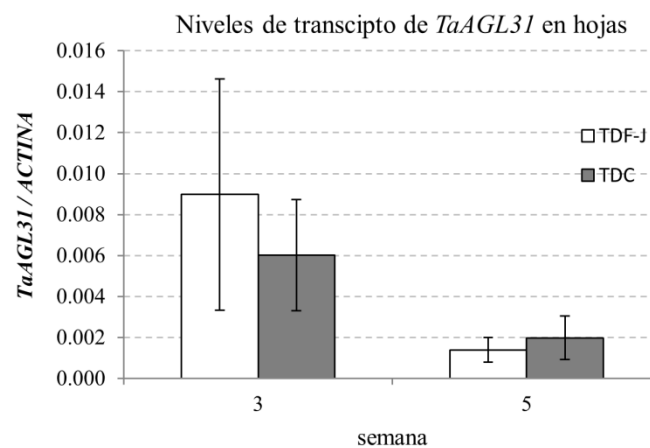
El gen Bradi4g40350 de *B. distachyon*, el cual muestra una identidad del 90% con el gen de trigo *AGAMOUS-like 31* (*TaAGL31*) es un factor de transcripción del tipo MADs-BOX y representa el mejor candidato para *VRN-D4* encontrado en esta región. En *Arabidopsis* *AGAMOUS* es un MADs-BOX gene relacionado con identidad floral. *TaAGL31* es 100% idéntico a la secuencia detectada en el cromosoma 5DS (IWGSC\_chr5DS\_ab\_k71\_2764238, <http://wheaturgi.versailles.inra.fr/Seq-Repository>), el mismo brazo cromosómico de *VRN-D4*. Usando BLAST reverso se confirmó que Bradi4g40350 es el gen más cercano a este gen de trigo.

<i>B. distachyon</i>	Arroz	Dominio proteico conservado [Identificador de dominio en NCBI]	Arabidopsis BLASTP Identificador en NCBI y valor E	Ortologo en trigo (URGI contig ID)
Bradi4g39630	Os12g13170	B_zip[cd12193], basic leucine zipper DNA-binding and multimerization region of GCN4 and related proteins; MFMR[pfam07777], G-box binding protein MFMR.	NP_850248.2; Basic region/leucine zipper transcription factor 16; E= $1e^{-120}$ .	IWGSC_chr5DL_ab_k71_4513163
Bradi4g40057	Not Found	Zinc finger, PHD-type (IPR001965), Zinc finger, RING/FYVE/PHD-type.	NP_200350; Enhanced downy mildew; E= $2e^{-63}$ .	IWGSC_chr5DL_ab_k71_4518453
Bradi4g40230	Not colinear	DUF3594 super family [cl13583], RING super family [cl15348], RING-finger domain, specialized type of Zn-finger.	XP_002875129.1; PHD finger family protein; E= $2e^{-89}$ .	No encontrado
Bradi4g40250	Os12g10660	BBOX super family [cl00034], B-Box-type zinc finger; zinc binding domain (CHC3H2); $3.56e^{-04}$ .	XP_002887747.1; zinc finger (B-box type) family protein; E= $3e^{-38}$ .	IWGSC_chr6DS_ab_k71_2071445
Bradi4g40270	Os12g10630	ZF-HD_dimer super family [cl04737]. homeo_ZF_HD super family[cl11752], homeobox domain, ZF-HD class.	NP_172896.1; AtHB31; DNA binding / transcription factor (AtHB31); E= $3e^{-48}$ .	No encontrado
Bradi4g40310	Not colinear	zf-TRAF super family[cl08341], TRAF-type zinc finger.	NP_187801.3; Prenyltransferase/zinc ion binding; E= $2e^{-84}$ .	IWGSC_chr7DL_ab_k71_3381171
Bradi4g40350	Os12g10540	MADS-box [cd00265] Type II subfamily of MADS. K-box super family [cl03234], K-box region.	CAA16753.1; floral homeotic protein agamous; E= $2e^{-67}$ .	IWGSC_chr5DS_ab_k71_2764238 (= TaAGL31= DQ512349)
Bradi4g40357	Os12g10520	MADS-box [cd00265] Type II subfamily of MADS.. K-box super family[cl03234], K-box region.	AAC49085.1; MADS-box protein AGL12; E= $2e^{-47}$ .	No encontrado
Bradi4g40540	Os12g09250	BRLZ[smart00338], basic region leucine zipper..	XP_002888232; predicted protein; E= $1e^{-68}$ .	IWGSC_chr5BS_ab_k71_2249326 (Pseudogene, not in 5AS & 5DS)
Bradi4g40517	Os12g09590	RRM_like_XS[cd12266], RNA recognition motif-like XS domain found in plants; zf-XS super family[cl04095], XS zinc finger domain.	XP_002872063; hypothetical protein ARALYDRAFT_489218; E= $1e^{-167}$ .	IWGSC_chr5DL_ab_k71_4509178

**Tabla 2.2. Selección de genes putativos anotados como factores de transcripción o genes relacionados con funciones en desarrollo. Solo se muestran los genes de trigo en la región candidata de *VRN-D4* que son colineales a las regiones del cromosoma 4 de *B. distachyon* y el cromosoma 12 de arroz.**

*TaAGL31* fue secuenciado en las líneas utilizadas para la población de mapeo [TDF-J y CS<sub>5402</sub>] y se usó uno de los SNPs encontrados para mapearlo en dicha población. *TaAGL31* fue mapeado completamente ligado a *VRN-D4* (Figura 2.2.A; Tabla 2.1).

También se secuenció la región promotora (1,404 bp río arriba del codón de inicio), el primer intrón, la región codificante y la región 3' UTR (438bp), tanto en línea portadora de *VRN-D4* (TDF) como en la línea isogénica que no posee *VRN-D4* (TDC). La región codificante de *TaAGL31* en estas dos líneas (KF761667, KF761666) es 100% idéntica a la secuencia del cultivar Chinese Spring disponible en Gen Bank (DQ512349). Finalmente se estudiaron los niveles de expresión de este gen en TDF y TDC mediante qRT-PCR a dos puntos diferentes del desarrollo (tres y cinco semanas, Figura 2.3), y no se encontraron diferencias en los niveles de expresión entre estas dos líneas.



**Figura 2.3. Comparación de los niveles de transcritos de *TaAGL31* en hojas de TDF (*Vrn-D4*) y TDC (*vrn-D4*) en platas de tres y cinco semanas de edad.** Los niveles de transcritos están expresados como veces-*ACTINA* y fueron calculados de acuerdo a la fórmula  $2^{(CT_{ACTINA} - CT_{TaAGL31})}$ . Las barras de error representan el error estándar de la media calculado a partir de 6 réplicas biológicas.

### 2.3.5 Epistásis entre *VRN-D4* y otros *loci* que controlan el requerimiento de vernalización

Se estudiaron las interacciones entre *VRN-D4* y otros *loci*, conocidos por controlar el requerimiento de vernalización (Tabla 2.3). El efecto de *VRN-D4*, los tres alelos de *VRN1* estudiados (*VRN-A1*, *VRN-B1* y *VRN-D1*) y *VRN-B3* fueron significativos ( $P < 0.0001$ ) en las cuatro poblaciones. Las interacciones entre *VRN-D4* y los otros *loci* estudiados fueron también significativas ( $P < 0.0001$ , Figura 2.4A-D). En los cuatro casos, *VRN-D4* mostró un pequeño efecto en tiempo a floración cuando los alelos

dominantes primaverales de los otros genes estuvieron presentes con respecto a cuando estuvieron ausentes (Figura 2.4A-D). Todas las plantas homocigotas para los alelos invernales de *vrn-D4* y *vrn-1* mostraron floración muy tardía en todas las poblaciones (medias ajustadas entre 119 y 130 días) (Figura 2.4A-D), confirmando que no hay genes adicionales segregando para el hábito de crecimiento primaveral en estas poblaciones. Esto fue también observado en la gran proporción de la variación en tiempo a floración ( $R^2 = 0.87\text{--}0.96$ ) explicada en cada población por el modelo de ANOVA factorial incluyendo los dos genes y sus respectivas interacciones.

**Tabla 2.3. Interacciones entre *VRN-D4* y otros alelos que regulan el requerimiento de vernalización.**

Se muestra la probabilidad de los efectos principales y las interacciones en un ANOVA factorial 2 x 3, la proporción de la variación explicada por el modelo ( $R^2$ ), el grado de dominancia y el efecto relativo de una combinación alélica de dosis simple.

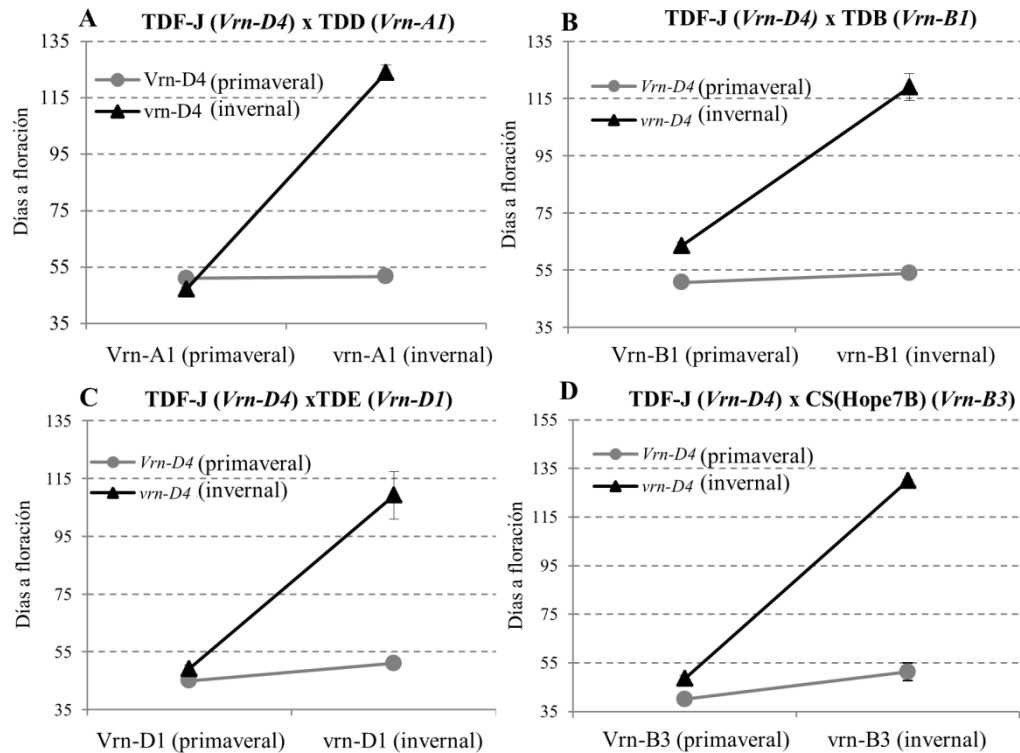
<i>VrnX</i>	<i>P</i> <i>VrnX</i> .	<i>P</i> <i>Vrn-D4</i>	<i>P</i> interacción	$R^2$	<i>VrnX</i> dominancia <sup>1</sup>	<i>Vrn-D4</i> dominancia <sup>1</sup>	<i>Vrn-D4 – VrnX</i> dosis simple <sup>2</sup>
<i>Vrn-A1</i>	<0.0001	<0.0001	<0.0001	0.96	95.9 %	53.7 %	18.5 d ***
<i>Vrn-B1</i>	<0.0001	<0.0001	<0.0001	0.91	55.6 %	30.7 %	1.0 d <sup>NS</sup>
<i>Vrn-D1</i>	<0.0001	<0.0001	<0.0001	0.89	78.2 %	42.9 %	14.3 d ***
<i>Vrn-B3</i>	<0.0001	<0.0001	<0.0001	0.87	71.0 %	31.3 %	18.9 d **

<sup>1</sup> El grado de dominancia fue calculado utilizando LS Means con la formula descrita en la sección de materiales y métodos.

<sup>2</sup> Los efectos de una dosis simple fueron calculados utilizando LS Means y la formula descrita en la sección de materiales y métodos. Los valores positivos indican que las plantas que llevan *VRN-D4* florecieron más tarde de las que llevaban el alelo de *VrnX* respectivo. \*\* =  $P < 0.001$  y \*\*\* =  $P < 0.0001$

El grado de dominancia (D) para el alelo de floración temprana de *VRN-D4* (30.7–53.7 %) fue más chico que el grado de dominancia de los diferentes alelos primaverales de *VRN1* (55.6–95.9 %, Tabla 2.3). De los alelos de *VRN1* estudiados, el alelo de *VRN-A1* para hábito de crecimiento primaveral mostró el grado más alto de dominancia, lo cual es consistente con previos estudios que muestran que una dosis de este alelo es suficiente para eliminar el requerimiento de vernalización. También se observó que *VRN-D4* tiene un efecto menor que *VRN-A1*, *VRN-D1* o *VRN-B3* en la aceleración de la floración. En un fondo genético donde todos los genes son homocigotas para hábito de crecimiento invernal, las plantas heterocigotas para *VRN-D4* florecieron en 14 y 19 días más tarde que aquellas heterocigotas para *VRN-A1*, *VRN-D1* o *VRN-B3* (Tabla 2.3). No

se encontraron diferencias significativas entre *VRN-D4* y *VRN-B1*, lo cual en conjunto con el valor similar de dominancia de estos dos genes sugiere que el alelo dominante de *VRN-D4* tiene un efecto similar en floración.



**Figura 2.4. Interacciones entre *VRN-D4* y otros alelos que regulan el requerimiento de vernalización.** Para simplificar la visualización solo se graficaron solo los valores de las plantas homocigotas para los diferentes alelos de vernalización segregando en cada población. (A) Interacciones entre *VRN-D4* y *VRN-A1*. (B) Interacciones entre *VRN-D4* y *VRN-B1*. (C) Interacciones entre *VRN-D4* y *VRN-D1*. (D) Interacciones entre *VRN-D4* y *VRN-B3*. Todas las interacciones son significativas ( $P < 0.0001$ , basado en ANOVA  $2 \times 3$ ).

## 2.4 DISCUSIÓN

### 2.4.1 Mapeo de *VRN-D4*

En el mapa de alta densidad presentado en este estudio, se separó a *VRN-D4* de cinco marcadores ligados a *VRN-D4* en previos mapas (Yoshida et al. 2010), confirmando la alta resolución de este nuevo mapa. Adicionalmente al número de recombinantes, este nuevo mapa provee 19 nuevos marcadores basados en secuencias codificantes en la

región centromérica del cromosoma 5D, lo cual facilitó las comparaciones de las regiones colineares con los genomas de *B. distachyon* y arroz.

El mapeo de alta densidad de *VRN-D4* y su eventual clonado posicional enfrentan dos desafíos principales. El primero es el limitado polimorfismo en el genoma D de trigo hexaploide, particularmente en regiones centroméricas (Dubcovsky y Dvorak 2007; Akhunov et al. 2010) y la limitada recombinación en la mitad proximal de los cromosomas de trigo (Luo et al. 2013; Lukaszewski y Curtis 1993). El primer problema fue resuelto en este estudio, utilizando para la población una línea (CS<sub>5402</sub>) que posee un cromosoma altamente polimórfico, perteneciente a *Ae. tauschii*. Se encontraron 120 SNPs en 24.4Kb secuenciados en los cromosomas de *T. aestivum* y CS<sub>5402</sub> (~5 SNP/Kb), en cambio cuando se secuenciaron las mismas regiones en las líneas TDC y Hayakomugi, no se encontraron SNPs.

Desafortunadamente, el alto nivel de polimorfismo está asociado con una reducción en la frecuencia de recombinación de alrededor del 30% en esta población (Yoshida et al. 2010). Esta reducción en la recombinación está posiblemente asociada con la habilidad del gene *Ph1* (Pairing homeologous 1) de detectar los relativamente altos niveles de heterocigocidad (Yousafzai et al. 2010; Griffiths et al. 2006) entre los cromosomas 5D de *T. aestivum* y *Ae. tauschii* impidiendo su recombinación. Por ejemplo, se ha descrito que *Ph1* está asociado con una reducción en la recombinación entre el genoma A de *T. aestivum* y el genoma A de *T. monococcum* (Dubcovsky et al. 1995) los que poseen en promedio ~12 SNPs/Kb en regiones de intrones (Dubcovsky y Dvorak 2007). Una estrategia posible para superar este obstáculo sería la utilización de un fondo genético con la mutación de pérdida de función de *Ph1*, sin embargo, incluso si los niveles de frecuencia de recombinación pueden ser restaurados a niveles normales, la baja recombinación en regiones centroméricas que es comúnmente observada continuará complicando los esfuerzos para disectar la región de *VRN-D4* por recombinación.

#### **2.4.2 Genética comparativa de la región candidata para *VRN-D4* en *B. distachyon* y arroz**

Los previos mapas de *VRN-D4* fueron desarrollados principalmente mediante la utilización de marcadores microsatélites (SSR) limitando las comparaciones con



genomas de otras especies (Yoshida et al. 2010). Los 20 marcadores basados en secuencias codificantes en la región centromérica del cromosoma 5D fueron mapeados en siete grupos que muestran buena colinearidad con la región del cromosoma 4 de *B. distachyon* y el cromosoma 12 de arroz (Figura 2.2). La mayoría de estos *loci* en *B. distachyon* fueron colineales con arroz, con la excepción del locus Bradi4g43750. Los únicos dos genes de trigo que no son colineales están localizados muy cerca y son colineales ni en *B. distachyon* o arroz (CJ521028 y FT-like). Los ortólogos de estos dos genes se encuentran a 0.06 Mb de distancia en el cromosoma 4 de *B. distachyon*.

Una interesante excepción de la buena colinearidad relativa observada entre los tres cromosomas de trigo, *B. distachyon* y arroz en la región de *VRN-D4* es la diferente posición del centrómero del cromosoma 5D. Los dos marcadores flanqueantes del centrómero en trigo (*TaAGL31* en el brazo corto y BE403761 en el brazo largo) delimitaron una región entre 44.74 y 45.12 Mb del cromosoma 4 de *B. distachyon* (Figura 2.2.B). Esta región se encuentra a más de 20Mb de distancia del centrómero activo del cromosoma 4 de *B. distachyon* [22–23 Mb, Qi et al. (2010)]. Esta región del centrómero activo del cromosoma 5D de trigo parece corresponder con un centrómero inactivo de *B. distachyon*. Utilizando sondas pericentroméricas marcadas del cromosoma 5D de trigo se observó que estas se localizan en el brazo largo del cromosoma 4 de *B. distachyon* en las posiciones 44–46 Mb (Qi et al. 2010). La secuenciación del genoma de *B. distachyon* confirmó que el centrómero activo del cromosoma 4 es homólogo al centrómero del cromosoma 4 de trigo y al del cromosoma 11 de arroz (International Brachypodium Initiative 2010). Los cromosomas 11 y 12 de arroz son el resultado de una duplicación ancestral que precede la divergencia de la mayoría de las subfamilias de los pastos (Paterson et al. 2004). El centrómero funcional del cromosoma 12 de arroz está delimitado por los marcadores Os12g10540 (5.58 Mb) y Os12g08260 (4.20 Mb) (Figura 2.2.C), lejos de la región colinear del centrómero del cromosoma 5D de trigo.

En resumen estos resultados indican que el centrómero del cromosoma 5D de trigo está en una localización distinta de los centrómeros del cromosoma 12 de arroz y del cromosoma 4 de *B. distachyon*.

### 2.4.3 Análisis de la región candidata de *VRN-D4*

*TaVIL-D1* fue previamente sugerido como un candidato para *VRN-D4* basado en su función homóloga en *Arabidopsis*, su localización centromérica en el cromosoma 5D y su inducción de la transcripción por vernalización (Fu et al. 2007). Los resultados del mapeo fino de *VRN-D4* aquí presentados concluyen que *TaVIL-D1* no es *VRN-D4*. Mediante dos eventos independientes de recombinación se demostró que *TaVIL-D1* es proximal a *VRN-D4*. Además, no se encontraron polimorfismos en las regiones codificante y regulatorias entre las diferentes líneas con distinto hábito de crecimiento estudiadas. Para analizar posible genes candidatos alternativos, se exploró la región colinear a *VRN-D4* en *B. distachyon* (The International Brachypodium Initiative 2010). Esta especie fue seleccionada dentro de las que poseen un genoma completamente secuenciado por su cercanía en relación evolutiva con trigo (Draper et al. 2001). La región colinear a *VRN-D4* en el cromosoma 4 de *B. distachyon* tiene una longitud de aproximadamente 1Mb e incluye 127 genes putativos. De acuerdo a la secuencia codificante de estos genes y su similitud estructural a nivel de proteína con otros genes y proteínas de función conocida la mayoría de estos ha sido anotado como relacionado con funciones estructurales o metabólicas, por lo que representan candidatos muy poco probables para *VRN-D4*. Sin embargo, diez genes en esta región fueron anotados como factores de transcripción o con función relacionada a procesos del desarrollo (Tabla 2.2). Mediante un análisis detallado de estos genes se mostró que únicamente un único gen representa un candidato probable para *VRN-D4* (Bradi4g40350). Este gen en trigo es denominado *AGL31* y presenta una alta similitud con el gen de *Arabidopsis* *AGAMOUS*. Debido a que *VRN1* es un factor de transcripción del tipo MAD-box y juega un papel importante en la regulación de la floración (Yan et al. 2003), inicialmente postulamos que *TaAGL31* podría jugar un papel similar. Sin embargo, la ausencia de polimorfismos detectados entre líneas con diferentes alelos de *VRN-D4* (TDF y TDC) y la ausencia de diferencias en los niveles de transcripción hacen que *TaAGL31* sea un candidato poco probable para *VRN-D4*. Aunque no se pueden descartar variaciones en las regiones regulatorias de este gen no abordadas en este estudio o diferencias en los niveles de expresión fuera de los puntos analizados, los resultados presentados aquí proveen poca evidencia para respaldar la hipótesis de *TaAGL31* como *VRN-D4*.

Aunque no podemos descartar la posibilidad de que algún otro gen presente en la región colinear de *B. distachyon* sea el verdadero homólogo de *VRN-D4*, nos inclinamos hacia la hipótesis de que *VRN-D4* se originó después de la divergencia evolutiva en *B. distachyon* y trigo. Esta hipótesis es sustentada en la ausencia de variación natural para el requerimiento de vernalización en las regiones colineares de los cromosomas 5A y 5B. Además, la limitada distribución geográfica del alelo dominante de *VRN-D4* (Iwaki et al. 2000, 2001) argumenta también a favor de un origen relativamente reciente.

#### 2.4.4 Interacciones epistáticas de *VRN-D4*

En un estudio previo se demostró que en trigo, la presencia de *VRN-D4* en la línea TDF estaba asociada con un incremento en los niveles de transcritos de *VRN1* y *VRN3* (*TaFT1*) y una disminución en los niveles de transcripción de *VRN2* (Yoshida et al. 2010). La activación del circuito de retroalimentación positiva que forman estos tres genes resulta en una acumulación de los niveles de transcripción de *VRN3* hasta llegar a niveles muy altos (>diez veces los niveles de *ACTINA*). La proteína producida por *VRN3* (*TaFT1*) es transportada al meristema apical donde induce *VRN1* (y probablemente sus parálogos *FUL2* y *FUL3*) (Chen y Dubcovsky 2012) donde inicia la transición del estado de meristema vegetativo a reproductivo.

Basados en los resultados previos de expresión génica, hipotetizamos que *VRN-D4* debería ser activo en las hojas y debería operar ‘rio arriba’ o ser parte del circuito de retroalimentación positiva que forman *VRN1/VRN2/VRN3*. Las interacciones epistáticas presentadas en este estudio soportan esta hipótesis ya que en la presencia de los alelos dominantes de *VRN1* (que eliminan o reducen el requerimiento de vernalización para activar el circuito de retroalimentación positiva) la adición de *VRN-D4* tiene un efecto pequeño en el tiempo a floración. Además, cuando *VRN-B3* está presente (alelo dominante de *FT1*), los altos niveles de *FT1* resultan en la regulación positiva de *VRN1* y la regulación negativa de *VRN2*, más allá de la presencia o no de *VRN-D4*.

En plantas con hábito de crecimiento invernal el circuito de retroalimentación positiva no está activo. En este fondo genético, la presencia de *VRN-D4* es suficiente para activar este circuito, resultando en una floración más temprana que las líneas invernales. Debido a que los cambios en los perfiles de expresión de cualquiera de estos genes directamente afecta la expresión de los otros dos, actualmente es difícil predecir cuál de estos genes es el blanco principal de *VRN-D4*.

## 2.5 CONCLUSIÓN

En esta primera etapa se comprobó la presencia de *VRN-D4* en el cromosoma 5D como había sido previamente reportado (Yoshida et al. 2010), pero también se logró avanzar en el desarrollo de un mapa genético detallado, donde finalmente se pudo descartar la parte proximal del brazo largo permitiendo focalizar el análisis a la parte proximal del brazo corto de este cromosoma.

Un análisis genómico comparativo, no reveló potenciales genes candidatos apoyando la hipótesis que *VRN-D4* es el resultado de un gen que no se encuentra conservado en los otros genomas de trigo, y por lo tanto tampoco en especies relacionadas.

Si bien el uso de un cromosoma D de *Ae. tauschii* en la línea CS<sub>5402</sub> (población de mapeo TDFxCS<sub>5402</sub>) facilitó la identificación de polimorfismos en la región de *VRN-D4*, también resultó en una reducción de la recombinación en esta región que dificulta la separación de *VRN-D4* de la zona centromérica.. La publicación de este estudio completa el primer objetivo de esta tesis: el desarrollo de un mapa genético detallado de la región de *VRN-D4*. Este estudio también presenta una descripción detallada de interacciones epistáticas con los otros genes de vernalización. Estos resultados, junto con la caracterización de los patrones de expresión de *VRN1/VRN2/VRN3* permiten hipotetizar que *VRN-D4* opera como parte del circuito *VRN1/VRN2/VRN3* o como regulador de uno de estos genes.

## CAPÍTULO

## 3

## Identificación, validación y caracterización del gen candidato para *VRN-D4*

---

Nestor Kippes, Juan M. Debernardi, Hans Vasquez-Gross, Bala A. Akpinar, Budak Hikmet, Kenji Kato, Shiaoman Chao, Eduard Akhunov and Jorge Dubcovsky (2015) **Identification of the *VERNALIZATION 4* gene reveals the origin of spring growth habit in ancient wheats from South Asia.** Proc Natl Acad Sci USA 112 (39):E5401-E5410.

### 3.1 INTRODUCCIÓN

En el capítulo anterior se mostró que *VRN-D4* se encuentra localizado en la región centromérica del cromosoma 5D y la región candidata se delimitó a la región proximal del brazo corto. El hecho de que la frecuencia de recombinación sea muy baja en la población utilizada, demuestra que no es posible utilizar clonado posicional para aislar este gen por recombinación. Además también, el análisis de los genes candidatos en dicha región concluyó que no fue posible encontrar genes que puedan ser fácilmente identificables como candidatos para *VRN-D4* basados en su función predicha en el genoma de la especie modelo de los pastos *Brachypodium distachyon*.

Un ejemplo muy similar a esta situación es la que Xue W. et al (2008) enfrentaron para la identificación del gen de arroz *Ghd7* que controla el número de granos, floración y altura de planta (homologo a *VRN2* en trigo). Este gen fue inicialmente identificado en una región centromérica del cromosoma 7 de arroz, donde la recombinación es 37 veces menor a la promedio. El hecho de contar con un proyecto genoma completo facilitó la identificación de los genes predichos en esa región (450 genes) y ,aunque la falta de recombinación impidió la reducción de esa lista de genes candidatos, los autores utilizaron la predicción de los dominios de *Ghd7* (que posee dominios CCT como el gen CO, COL (CO like) o TOC1) para identificarlo como posible candidato y continuar con el estudio de este gen, que finalizó con la descripción de *Ghd7* como gen que controla estas características.

En este capítulo se describen las estrategias alternativas utilizadas a fin de identificar genes candidatos para *VRN-D4*, así como los métodos de validación y análisis de distribución geográfica.

### 3.2 MATERIALES Y MÉTODOS

#### 3.2.1 Estudio de expresión de genes mediante RT-QPCR

Ocho plantas por genotipo fueron cultivadas en cámara de cultivo (PGR15, CONVIRON, USA) con fotoperíodo largo (16h de luz) y condiciones no vernalizantes (18°C durante la noche y 20°C durante el día). La colecta de tejido (última hoja

totalmente desplegada) y la extracción de ARN se produjeron de la misma manera que fue esta detallada en la sección de Materiales y Métodos del Capítulo 2. Los cebadores utilizados fueron previamente publicados: *VRN1*, *VRN2*, *VRN3* (Yan et al. 2006 y Distelfeld et al. 2009a), los tres homólogos de *VRN1* (Loukoianov et al. 2005) y *ACTINA* (Uauy et al. 2006).

### 3.2.2 Búsqueda y análisis de mutaciones en la población de TILLING

Una población de 1,153 plantas de la línea portadora de *VRN-D4* obtenidas después del tratamiento con EMS 0.9% fue usada para buscar mutaciones mediante la técnica de *CeII* previamente descrita (Uauy et al. 2009). El primer par de cebadores utilizados (TILL-5DS-F1 y TILL-A-DS-R10, Tabla A.3) incluye el SNP A367C específico de *VRN-D4* y amplifica específicamente la región del exón 4 al 8 de este gen. Dos pares de cebadores adicionales, desarrollados previamente (Chen y Dubcovsky 2012), y específicos para *VRN-A1* fueron utilizados para amplificar dos regiones distintas de *VRN-D4*. Los cebadores V1A-5PRIME-F1 y V1-INT1R amplifican una región que incluye el exón 1, mientras que los cebadores VA1-3PRIME-F2 y VA1-3PRIME-R3 amplifican una región que cubre del exón 3 al 6. A pesar que estos cebadores amplificaron en simultáneo *VRN-D4* y *VRN-A1* las mutaciones fueron detectables mediante *CeII*. Para determinar cuál de estos dos genes poseía las mutaciones encontradas se secuenció el producto de PCR y caracterizó la mutación y el efecto de la misma, mediante predicción de cambio a nivel de proteína. Los efectos a nivel de proteína fueron analizados usando los algoritmos de predicción PROVEAN (Choi et al. 2012), SIFT (Ng et al. 2001) y PolyPhen-2 (Adzhubei et al. 2010) a fin de seleccionar aquellas mutaciones que pueden tener un mayor efecto sobre la funcionalidad de la proteína.

Debido a que *VRN-D4* se expresa en altos niveles desde estadíos tempranos del desarrollo, se extrajo ARN de la primera hoja completamente expandida de los mutantes de interés y se usó la secuenciación de la región codificante de *VRN-D4* a modo de corroborar que las mutaciones obtenidas se encontraron en los mismos transcritos que el SNP característico de *VRN-D4*. En aquellos mutantes donde se amplificaron en simultáneo *VRN-A1* y *VRN-D4* se usó el clonado del producto de amplificación de ADNc para secuenciar transcritos individuales.

Cebadores específicos para *VRN-D4*: Mediante los cebadores para amplificar específicamente *VRN-D4* se detectaron 21 mutaciones. Diez de estas mutaciones ocurrieron en regiones codificantes de *VRN-D4* y cuatro de estos cambios determinan cambios a nivel de proteína. Mediante los algoritmos de predicción de efecto se seleccionó la mutación E158K localizada en el dominio K-box ya que fue predicha como la de mayor efecto entre las mutaciones encontradas (Tabla 3.1). Los mismos algoritmos sugirieron un efecto limitado para las otras tres mutaciones (A180T, D184N y A219T) (Tabla 3.1). Basado en estos resultados las últimas tres mutaciones fueron descartadas y solamente el cambio E158K fue seleccionado para caracterización fenotípica.

Cebadores para *VRN-D4* y *VRN-A1*: mediante la búsqueda con los cebadores 5PRIME-F1 y V1-INT1R que cubren el prime exón se encontraron 19 mutantes, pero solo tres de estos ocurrieron en la región codificante del primer exón. Dos de estas mutaciones produjeron cambios de aminoácidos (E12K y E42K) pero, desafortunadamente ambas fueron identificadas en *VRN-A1*.

Con los cebadores VA1-3PRIME-F2 y VA1-3PRIME-R3 se detectaron siete mutantes entre la región del exón 3 al 6. Una de estas mutaciones fue predicha de producir un cambio de aminoácido (E129K) pero fue identificada en *VRN-A1*. Una segunda mutación detectada resulta en un cambio en el sitio dador de splicing del cuarto exón. La pérdida del sitio de splicing produce dos codones de STOP de la traducción localizado 8pb río debajo de la mutación. La introducción de este codón de STOP es predicha de truncar la proteína después de la posición 143 lo que elimina 30% del dominio K-box y más del 40% de la proteína. Para confirmar la presencia del codón prematuro de *STOP* predicho en los transcritos de *VRN-D4* se extrajo ARN de la primera hoja expandida de este mutante y de plantas TDF como control. Los transcritos amplificados de ADNc fueron clonados usando el sistema NEB® PCR Cloning (New England Biolabs, USA) y secuenciado mediante Sanger sequencing. La presencia de la mutación del sitio de splicing en conjunto con la retención del intrón correspondiente así como el SNP característico de *VRN-D4* fue corroborado en 16 clones independientes confirmando que esta mutación se corresponde con este gen. En contraste, todas las secuencias de 22 clones de la planta usada como control mostraron el splicing correcto con el SNP correspondiente a *VRN-D4*.



Análisis de conservación del sitio de *VRN-D4* mutando por TILLING: Se seleccionaron 268 proteínas homólogas a *VRN-D4* usando la secuencia de proteína de *VRN-D4* para una búsqueda de BLAST ([provean.jcvi.org](http://provean.jcvi.org)). Con estas secuencias se produjo un alineamiento múltiple usando el software MEGA 6 (Tamura et al. 2013) ([www.megasoftware.net](http://www.megasoftware.net)) y el resultado de este alineamiento fue usado para generar un weblogo de la secuencia proteica ([weblogo.berkeley.edu](http://weblogo.berkeley.edu)), en el cual cada letra representa los aminoácidos encontrados en esa determinada posición y el tamaño es proporcional a la frecuencia de ese particular aminoácido en el alineamiento múltiple.

### 3.2.3 Secuenciación de BACs y mapa físico

Se realizó una búsqueda en el mapa físico del brazo corto del cromosoma 5D de Chinese Spring, mediante el uso de tres pares de cebadores en distintas regiones de *VRN-D4* (Tabla A.3). Esta biblioteca de BACs fue desarrollada por el Instituto de Botánica Experimental (República Checa) como parte del proyecto internacional de genoma de trigo (IWGSC). Se identificó un BAC (TaeCsp5DShA\_0038\_M14) que es parte de dos contigs mayores (CTG87 y CTG61) (Tabla A.4). Se seleccionaron un total de 12 BACs de estos dos contigs para secuenciación. Se produjeron bibliotecas para secuenciación con Illumina usando el kit KAPA LTP Library Preparation (KAPA Biosystems, USA) con un tamaño promedio de inserto de 250 pb; también se produjeron cuatro bibliotecas adicionales con un tamaño de inserto de 400 pb para incrementar la calidad del ensamblado en esta región. Las bibliotecas fueron marcadas con distintos códigos de barras para su posterior identificación, mezcladas y secuenciadas en el Instituto Genómico de Beijing (BGI, Sacramento, CA, USA) usando lecturas de 100-pb con extremos solapados (Illumina Hi Seq2500).

El ensamblado de los BACs se produjo usando Abyss v1.5.2 (k=63) ([www.bcgsc.ca/platform/bioinfo/software/abyss](http://www.bcgsc.ca/platform/bioinfo/software/abyss)) y combinando las bibliotecas con tamaño de inserto de 250 y 400pb (la cobertura individual de cada BAC fue mayor a 100X). Debido a que se encontró un extenso solapamiento entre BACs, se obtuvo un mejor ensamblado cuando se analizaron grupos de cuatro BACs para el ensamblado en Abyss (k=63). Se generaron lecturas adicionales de los mismos BACs mediante secuenciación 454 que fueron usadas para completar gaps en el ensamblado obtenido mediante la secuenciación con Illumina. La secuenciación mediante la tecnología 454

fue realizada en un equipo GS FLX Titanium (Roche 454 Life Sciences, Branford, CT, USA) usando los materiales provistos por el fabricante para la construcción de las bibliotecas. El ensamblado obtenido de ~680Kb fue depositado en GenBank.

Los genes y elementos repetitivos fueron anotados usando la herramienta TriAnnot (<https://urgi.versailles.inra.fr/Tools/TriAnnot-pipeline>). Toda la región fue comparada con las secuencias presentes en las bases de datos 5AL y 5DS del IWGSC (<https://urgi.versailles.inra.fr/blast/>), con *T. urartu* (Ling et al. 2013) y con tres bases de datos distintas *Ae. tauschii* ([http://plants.ensembl.org/Aegilops\\_tauschii/](http://plants.ensembl.org/Aegilops_tauschii/), GCA\_000347335.1 (Jia et al. 2013), <http://aegilops.wheat.ucdavis.edu/ATGSP/>, (Luo et al. 2013), y [https://urgi.versailles.inra.fr/download/iwgsc/TGAC\\_WGS\\_assemblies\\_of\\_other\\_wheat\\_species](https://urgi.versailles.inra.fr/download/iwgsc/TGAC_WGS_assemblies_of_other_wheat_species), TGAC\_WGS\_tauschii\_v1). Los resultados fueron filtrados para conservar secuencias con una identidad mayor al 99% y longitud mayor a 500 pb. (Figura 3.6). Los resultados fueron convertidos a formato .gff y visualizados usando el programa IGV ([www.broadinstitute.org/igv/](http://www.broadinstitute.org/igv/)).

### 3.2.4 Marcadores para el sitio de inserción 5DS/5AL

Mediante los análisis de similitud con las secuencias pertenecientes al genoma A y al genoma D descrito anteriormente se identificó el primer borde de la inserción 88,344pb río arriba del codón de inicio de la traducción de *VRN-D4*. Río arriba de este sitio la secuencia del CS-dt5DS fue 99.3% idéntica al contig 5265.1 de *Ae. tauschii* y río abajo de este punto la secuencias es 99.6 % idéntica al contig IWGSC\_5AL\_2795346 de brazo cromosómico 5AL. Se diseñaron los cebadores VRND4-ins.F4 (CGTCTACAGAACCAGCTGAC) y VRND4-ins.R3 (GAAAATCAGGGTGTGACAAA) situados a ambos lados del punto de inserción para confirmar la existencia física de este evento. Este diseño de cebadores de PCR produjo una banda de amplificación de un 1440pb de tamaño únicamente en las líneas portadoras de *VRN-D4*. Como control se incluyó otro conjunto de cebadores que amplifican un gen localizado en el brazo largo del cromosoma 5D (BJ315664). Estos cebadores BJ315664-F (ACTTGCGTTCCATGTTTCACTTGTAGG) y BJ315664-R (GCGCTCCTGGCAGCAGTCTC) producen un producto de amplificación de 687 pb, que fue identificado en todas la líneas utilizadas.

El segundo borde de la inserción 5DS/5AL fue identificado en un región de 1.7 Kbp localizado entre 201,341 y 203,027 pb río debajo de *VRN-D4*. Las secuencias flanqueantes fueron encontradas, río arriba de la inserción, 98.7% idénticas al contig 2801682 del cromosoma 5AL y 99.8% idénticas al contig 280242 (TGAC\_WGS\_tauschii\_v1) de *Ae. tauschii* río abajo del punto de inserción. Se diseñaron cebadores a ambos lados de la inserción (VRND4-ins2.F1 TATCTACACCCAGCGAGAGA y VRND-ins2.R1-GGCCTACCAAAAACAAAAGT) para amplificar un fragmento de PCR de 1283 pb de longitud, y se usó un gen localizado en el cromosoma 5D como control (BE606654-F-ACCGACAAGAACTCCTAGAT y BE606654-R-CAAAAGCATCGCAGAGAAACAC). Este control amplificó un fragmento de 534pb en todas las líneas utilizadas. Las reacciones de PCR fueron llevadas a cabo en un volumen final de 20 µl conteniendo 10 mM Tris-HCl, PH 8.3, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 0.2 mM dNTP, 0.5 µM de cada primer para la inserción y 0.05 µM de cada primer control, 50–100 ng de ADN genómico, y 0.5 U Taq DNA polimerasa. Los ciclos de PCR fueron llevado a cabo como se describe a continuación: 5 min a 94°C, 8 ciclos en un gradiente de temperatura comenzando a 65°C (-0.5°C cada ciclo), seguido de 32 ciclos como se detalla a continuación: 30seg a 94 °C, 30seg a 61 °C, y 1.30 min a 72 °C, con una extensión final por 7 min a 72 °C.

### 3.2.5 RNA-EMSA

Las sondas de ARN biotiniladas en el extremo 3' utilizadas fueron adquiridas de SIGMA-ALDRICH (St. Louis, MO USA). Para los ensayos de unión se utilizaron las proteínas purificadas RNase-free GST-TaGRP2 y GST (control). GST-TaGRP2 fue clonada en el vector pGEX-6P-1, expresadas en *E. coli*, y purificadas usando la columna GST SpinTrap Purification Module (GE Healthcare) (Figura 3.12.A). El ensayo de movilidad electroforética de RNA (EMSA) fue llevado a cabo de acuerdo con las recomendaciones del fabricante utilizando el LightShift® Chemiluminescent RNA EMSA Kit (Pierce, USA), se incluyó además DMSO 5% en la reacción de unión para evitar uniones no específicas. Para reducir variaciones en la calidad de la proteína entre reacciones se realizó una mezcla principal con todos los componentes de la reacción sin incluir las sondas de ARN. También se incluyó un paso de pre-incubación de las sondas de ARN por 3 minutos a 80°C para favorecer la completa desnaturalización de las estructuras del ARN. Las muestras fueron incubadas a temperatura ambiente para

permitir re-naturalización antes de agregarlas a la reacción. Un experimento separado fue hecho transfiriendo las muestras de temperatura ambiente e incubando a 4°C, para a continuación llevar a cabo la reacción de unión a esta temperatura y de esta forma simular condiciones de vernalización. Las secuencias de las sondas utilizadas están listadas en la Tabla A.3. La estructura de las sondas utilizadas y de las secuencias flanqueantes a RIP-3 fueron generadas utilizando la secuencia de *T. monococum* con el programa RNAfold (Gruber et al. 2008) usando los parámetros por defecto.

### 3.2.6 Análisis de estructura de la población

Se utilizaron 100 líneas de *T. aestivum* pertenecientes a seis subespecies diferentes: *T. aestivum* ssp. *aestivum* (10 entradas primaverales y 10 invernales), *T. aestivum* ssp. *compactum* (17 entradas), *T. aestivum* ssp. *macha* (15 entradas), *T. aestivum* ssp. *spelta* (15 entradas), y *T. aestivum* ssp. *sphaerococcum* (33 entradas) (Tabla A.5). Para determinar la estructura de la población se utilizaron un total de 16371 marcadores SNP polimórficos usando un modelo de agrupamiento Bayesiano (Pritchard et al. 2000) implementado en el programa STRUCTURE v.2.3.4 (Pritchard et al. 2000). El número de subpoblaciones (K) analizado varió de 2 a 8. Cada valor de K fue simulado por separado, empezando con un período de 5000 iteraciones de entrenamiento, seguido de 10000 réplicas para la fase de muestreo. Para detectar el número de subpoblaciones que mejor se ajusta a la muestra, se graficó la probabilidad media de los valores de K usando STRUCTURE HARVESTER v 0.6.93 (Earl et al. 2012). La matriz de correlación se calculó usando los 16371 marcadores en el programa TASSEL 4.3.10 ([www.maizegenetics.net](http://www.maizegenetics.net)). La matriz fue representada en un gráfico de gradiente de colores donde cada color representa un valor en R ([www.r-project.org](http://www.r-project.org)). Los cálculos del índice de diferenciación (Fst) fueron calculados con el programa Arlequin versión 3.0 (Excoffier et al. 2005)

### 3.2.7 Diversidad genética

El índice de contenido polimórfico (PIC) (Anderson et al. 1993) fue calculado por separado para cada una de las cinco subespecies, de acuerdo a la siguiente fórmula:

$$PIC = 1 - \sum_i^n p_i^2$$

donde  $p_i$  es la frecuencia del alelo  $i^{th}$ .

La diversidad genética fue estimada utilizando la fórmula de Weir (1996) la cual es equivalente a un promedio del PIC.

$$D = 1 - \frac{1}{L} \sum_l \sum_i p_i^2$$

donde  $p$  es la frecuencia del alelo  $i$  para el locus  $l$  y  $L$  es el número de *loci*. Los valores de PIC para cada subespecie fueron promediados para cada genoma (Tabla 3.4), cromosomas (Figura 3.15), y regiones cromosómica (Figura 3.17.E). Para reducir la posible introducción de sesgos en el análisis, los valores de diversidad genética por cromosoma y por región cromosómica fueron normalizados dividiéndolos por la diversidad genética de la correspondiente subespecie y genoma (mostrados en Tabla 3.4). Esta relación es referida como “diversidad genética normalizada”. Los valores mostrados en la Tabla 3.4 no pueden ser usados para comparar niveles relativos de diversidad genética entre subespecies ya que mucho de los SNP utilizados en el diseño de éste fueron específicamente seleccionados para una alta frecuencia de alelos menores en *T. aestivum* ssp. *aestivum* (Wang et al. 2014). Estos valores fueron calculados solamente para normalizar la diversidad genética por cromosoma y región cromosómica. Estos análisis fueron hechos con un subconjunto de 14,236 marcadores SNP donde la ubicación cromosómica es conocida.

### 3.3 RESULTADOS

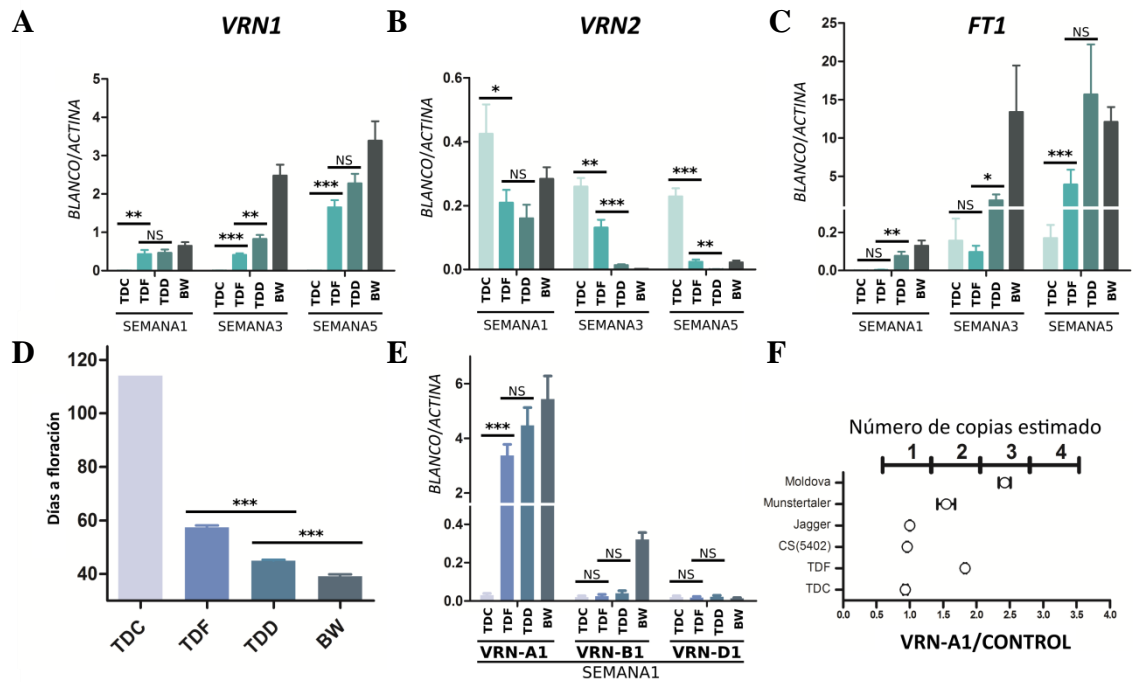
#### 3.3.1 Caracterización de la línea portadora de *VRN-D4*

En un estudio previo se mostró que *VRN-D4* estaba asociado con aumentos en los niveles de expresión de *VRN1* y *FT1* (*VRN3*) y con una disminución de los niveles de expresión de *VRN2* en un estadio avanzado de la etapa reproductiva (Yoshida et al. 2010). Basado en este resultado y en los resultados de interacciones epistáticas detallados en el capítulo anterior se propuso que *VRN-D4* opera río arriba o es parte del circuito que involucra a *VRN1*, *VRN2* y *FT1*. Para estudiar cuál de estos tres genes es el

blanco de *VRN-D4* realizamos un estudio de expresión génica de estos genes durante el desarrollo en líneas isogénicas portadoras y no portadoras de *VRN-D4* comenzando el análisis desde la primera hoja completamente expandida hasta la espigazón.

La línea Triple Dirk C (TDC, hábito de crecimiento invernal, *vrn-D4*) mostró bajos niveles de *VRN1* y *FT1* y altos niveles del represor *VRN2* en los tres puntos estudiados (semana uno, tres y cinco). En contraste la línea Triple Dirk F (TDF, hábito de crecimiento primaveral, *Vrn-D4*) mostró altos niveles de *VRN1* y *FT*, y bajos niveles de *VRN2* a semana cinco (Figura 3.1.A-C) confirmando los resultados previos (Yoshida et al. 2010). En el punto más temprano del desarrollo estudiado, las diferencias en la expresión de *VRN1* entre TDF y TDC fueron altamente significativas ( $P = 0.002$ , aumento de 38 veces, Figura 3.1.A), mientras que las diferencias en *VRN2* fueron marginalmente significativas ( $P = 0.05$ , disminución 2 veces, Figura 3.1.B). Las diferencias en los niveles de transcripción de *FT1* entre TDF y TDC fueron detectables solamente en el último punto de muestreo (semana cinco, Figura 3.1.C). Estos resultados sugieren que, entre los genes estudiados, *VRN1* es el blanco más temprano de inducción por *VRN-D4*.

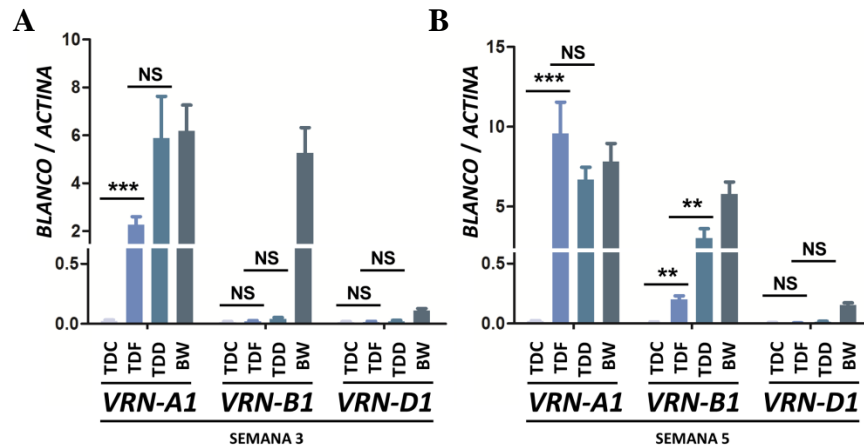
En ausencia de vernalización, TDF floreció mucho antes que TDC pero aún 12.6 días más tarde que Triple Dirk D (TDD, alelo primaveral *Vrn-A1a* (Yan et al. 2004)) y 18.3 días más tarde que Bobwhite (BW, el cual posee alelos primaverales en los tres homólogos de *VRN1*) (Figura 3.1.D). La floración más tardía en relación a TDD y BW estuvo asociada con niveles de transcripción de *VRN1* y *FT* más bajos, y más altos niveles de *VRN2* (Figura 3.1.A-C, semana 3).



**Figura 3.1. Caracterización de la línea TDF (*Vrn-D4*).** Niveles de transcritos de (A) *VRN-1* (todos los homólogos combinados), (B) *FT1* y (C) *VRN-2*. (D) Efecto en tiempo a floración de *VRN-D4*. El experimento fue terminado 114 días después de la siembra. (E) Niveles de transcripción de los tres diferentes homólogos de *VRN*. (F) Número haploide de copias de *VRN-A1*. Jagger, Munstertaler y Moldova fueron incluidos como controles para 1, 2 y 3 copias de *VRN-A1* respectivamente (Zhu et al. 2014). (A-E) Las medias fueron calculadas a partir de ocho replicas biológicas. (F) Media calculadas a partir de cuatro replicas biológicas. Las barras de error en todos los gráficos representan el error estándar de la media (SEM). TDC y CS<sub>5402</sub>: invernial, sin alelos de hábito primaveral. TDF: primaveral, *Vrn-D4*. TDD= primaveral, *Vrn-A1a*. BW (Bobwhite): primaveral, *Vrn-A1*, *Vrn-B1*, *Vrn-D1*. NS= no significativo, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , y \*\*\* =  $P < 0.001$ ).

Sorpresivamente, cuando se estudió la expresión de los tres homólogos de *VRN1* por separado (*VRN-A1*, *VRN-B1* y *VRN-D1*), se encontró que solamente *VRN-A1* fue regulado positivamente en la línea TDF (portadora de *VRN-D4*). Los niveles de transcripción de *VRN-A1* se encontraron aumentados desde la semana uno hasta semana cinco y fueron similares a los observados en la línea TDD, portadora del alelo primaveral en *VRN-A1* (Figura 3.1.E). En contraste, los niveles de *VRN-B1* y *VRN-D1* no fueron detectables en TDF en los dos primeros puntos analizados (semanas 1 y 3) (Figura 3.2). Los niveles de expresión de *VRN-B1* y *VRN-D1* si fueron detectable en la línea Bobwhite después de la primera semana (Figura 3.2). Estos resultados sugieren que el homólogo del genoma A de *VRN1* (*VRN-A1*) es un blanco específico de *VRN-D4*

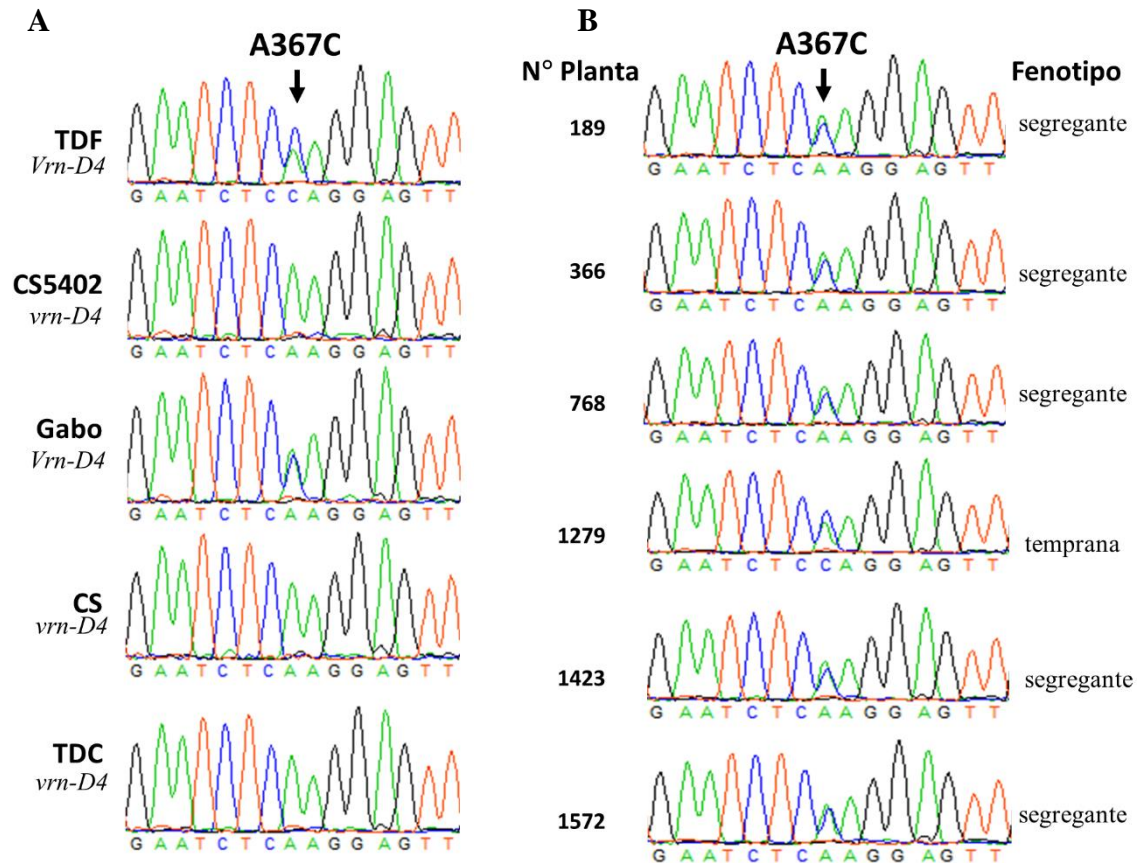
o alternatively los niveles elevados de *VRN-A1* son el resultado de copias adicionales de este gen.



**Figura 3.2. Niveles de expresión en hoja de los tres homólogos de *VRN1*.** (A) Plantas de tres semanas de edad (B) Plantas de cinco semanas de edad. Los resultados a semana 1, que forman parte del mismo experimento, son presentados en la Figura 3.1. Las barras representan la media calculada de ocho replicaciones biológicas y el error representa el error estándar de la media (SEM). El hábito de crecimiento y los alelos presentes se describen a continuación: TDC (invernal, *vrn-A1 vrn-B1 vrn-D1 vrn-D4*), TDF (primaveral, *vrn-A1 vrn-B1 vrn-D1 Vrn-D4*), TDD (primaveral, *Vrn-A1 vrn-B1 vrn-D1 vrn-D4*), y BW (Bobwhite, primaveral, *Vrn-A1 Vrn-B1 Vrn-D1 vrn-D4*). NS = no significativo, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , y \*\*\* =  $P < 0.001$ .

Utilizando la técnica de Taqman® previamente descrita para analizar el número de copias de *VRN-A1* (Diaz et al. 2012) se observó que las líneas portadoras del alelo recesivo de *VRN-D4*, TDC y CS<sub>5402</sub>, poseen un número haploide de copias de *VRN-A1* igual a uno y, sorpresivamente, TDF la línea portadora de alelo dominante de *VRN-D4* presentó un número de copias igual a dos (Figura 3.1.F). A continuación se secuenció la región codificante de *VRN-A1* a partir de ADNc producido mediante la extracción de ARN de hoja en TDF y también en la línea Gabo, que fue la fuente de *VRN-D4* utilizada por Sears en 1978 para introgresar *VRN-D4* en el fondo genético Triple Dirk y producir la línea isogénica TDF. Se encontró que únicamente las líneas portadoras del alelo dominante de *VRN-D4* poseían un cambio de nucleótido en la posición 367 de la secuencia codificante.

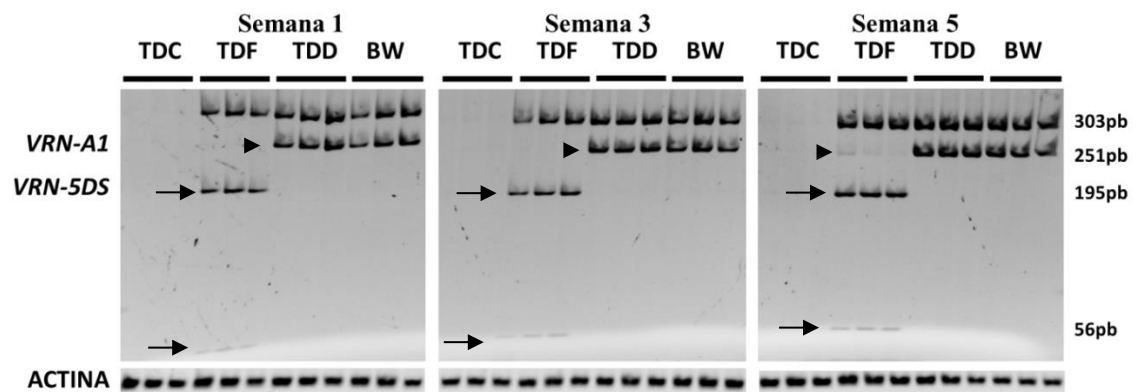




**Figura 3.3. Detección del SNP A367C de *VRN-D4*.** (A) Cromatogramas de secuenciación mostrando el polimorfismo A367C en las líneas portadoras del alelo dominante *Vrn-D4* (TDF y Gabo) y su ausencia en las líneas portadoras del alelo recesivo de *vrn-D4* (CS<sub>5402</sub>, CS: Chinese Spring y TDC). (B) Cromatogramas mostrando el polimorfismo A367C en los recombinantes críticos de la población TDF x CS<sub>5402</sub> detallada en el Capítulo 2. En los estudios de progenie realizados para cada individuo recombinante se estableció si dicha progenie segrega para el hábito de crecimiento primaveral (segregante). Cuando todos los individuos de esa progenie florecieron sin diferencias significativas entre ellos y de forma similar al padre portador de *VRN-D4* se los considera no segregantes y de floración temprana (temprana). No se encontraron recombinantes donde todos los individuos de la progenie tuvieron una floración tardía.

Las líneas con el alelo recesivo de *VRN-D4* presentaron una Adenina en la posición 367 (TDC, CS<sub>5402</sub> y Chinese Spring), que se corresponde con las secuencias disponibles conocidas de *VRN-A1*, por otro lado TDF y Gabo presentaron un cambio de Adenina por Citosina (A367C) que hasta la actualidad no había sido descrito (Figura 3.3.A). Este SNP fue estudiado en la población de mapeo fino descrita en el Capítulo 1, y fue encontrado completamente ligado con *VRN-D4* en el brazo corto del cromosoma 5D. Las seis plantas donde se encontraron los eventos de recombinación más cercanos a *VRN-D4* son mostradas en la Figura 3.3.B.

Para estudiar la expresión de esta copia adicional de *VRN-A1* localizada en el brazo corto del cromosoma 5D, se utilizó el sitio polimórfico de corte en las muestras de ADNc con la enzima de restricción *BstNI* que permite diferenciar a esta de la copia presente en la ubicación ancestral de *VRN-A1* en el cromosoma 5A (brazo largo). Las líneas portadoras de alelos dominantes primaverales para *VRN-A1* TDD y BW fueron incluidas como controles y en estas fue posible detectar expresión correspondiente a *VRN-A1* (variante ‘A’ del SNP 367) en el cromosoma 5AL visible como dos bandas de 303 y 251 pb.



**Figure 3.4. Expresión de los parálogos de *VRN1* del brazo cromosómico 5AL (*VRN-A1*) y 5DS (temporalmente denominado *VRN-5DS*). RT-PCR con cebadores que amplifican ambos genes (*VRN-A1* y *VRN-5DS*). Los dos transcritos pueden ser identificados mediante el polimorfismo A367C, el cual genera un sitio adicional de corte con la enzima de restricción *BstNI* en la secuencia de *VRN-5DS*. Luego de la digestión con *BstNI*, el fragmento de 251-pb detectado en *VRN-A1* (señalado con una cabeza de flecha) es cortado en dos fragmentos de 195 pb y 56 pb en *VRN-5DS* (flechas). En la línea TDF, donde *VRN-5DS* y *vrn-A1* están presentes, solamente los transcritos de *VRN-5DS* (variante ‘C’ del SNP 367) son detectados hasta la tercer semana. A las cinco semanas, una banda tenue correspondiente a *VRN-A1* también es detectada en TDF (variante ‘A’ del SNP 367). TDC tiene un hábito de crecimiento invernal y por lo tanto no expresa *VRN1*. Se incluyeron tres réplicas biológicas para cada genotipo. *ACTINA* (Yan et al. 2003) fue usado como control de carga de la amplificación por PCR.**

La copia del cromosoma 5D fue detectable únicamente en la línea portadora de *VRN-D4* TDF desde la semana uno y es identificable como tres bandas de 303, 195 y 56 pb respectivamente (variante ‘C’ del SNP 367). La expresión de esta copia fue detectable desde la semana uno del desarrollo mientras que la proveniente del cromosoma 5AL fue apenas detectable en TDF en el último punto estudiado (Figura 3.4 fragmento de

251pb). La línea invernal TDC fue usada como control negativo y no mostró expresión de *VRN1*, resultado característico de plantas en estado vegetativo.

### 3.3.2 Validación de *VRN-D4* usando mutaciones inducidas

La tecnología denominada TILLING (por Targeting Local Lesions IN Genomes), es una herramienta valiosa para estudios de genética funcional que combina mutagénesis química con un relevamiento a gran escala de mutaciones. El laboratorio del Dr. Dubcovsky cuenta con poblaciones mutantes de trigo hexaploide y tetraploide, compuestas por 1536 líneas cada una. Estas poblaciones se obtuvieron a partir de un tratamiento mutagénico con Etil Metano Sulfonato (EMS) y presentan una densidad promedio de 38 y 51 mutaciones por kb, respectivamente. Para facilitar un screening de alto rendimiento, estos ADNs se encuentran agrupados en conjuntos de 4. Para identificar mutantes en los genes de interés, se diseñan cebadores genoma-específico, que se utilizan para amplificar (mediante técnica de PCR) regiones de cada uno de estos genes, en la muestra conjunta. Los productos de PCR son desnaturalizados y re-hibridados, lo que permite la formación de heteroduplex (en el caso de que haya una mutación). Esas muestras son luego digeridas con una endonucleasa que reconoce y corta apareamientos incorrectos de un único nucleótido, y los fragmentos son visualizados en geles de poliacrilamida. Para los grupos que se detectan dos o más fragmentos, se secuencian los productos de PCR individuales a fin de determinar la mutación exacta que estos portan. Mediante este proceso, se han identificado numerosas líneas tetraploides y hexaploides las que han sido de gran utilidad para caracterizar genes importantes en las vías de regulación de la floración, resistencia a enfermedades y calidad panadera, por citar algunos. Debido a que ninguna de las poblaciones hexaploides actualmente disponibles es portadora de *VRN-D4*, para comprobar si la copia adicional de *VRN-A1* localizada en el brazo corto del cromosoma 5D es *VRN-D4* se produjo una nueva población de TILLING mediante la utilización de EMS a partir de la línea portadora de *VRN-D4* TDF. En total se obtuvieron 1,153 mutantes que fueron analizadas para mutaciones en *VRN-A1-5DS* usando la técnica de *Cell* (Uauy et al. 2009). Los detalles de cómo se realizó la búsqueda de mutantes y los cebadores utilizados se detalla en la sección de Materiales y Métodos de este capítulo. En total se

identificaron cuatro mutaciones no-sinónimas y un mutante de splicing en *VRN-A1-5DS*. Basado en la predicción del efecto de estas mutaciones a nivel de proteína (Tabla 3.1) se seleccionaron dos mutaciones para caracterización fenotípica. La primera mutación analizada genera un cambio de Acido glutámico por Lisina en la posición 158 de la proteína (E158K, Figura 3.5). Aunque este es un cambio de aminoácido tolerado (BLOSUM 62 score= 1), esta posición parece ser altamente conservada como se observa en el alineamiento de las 268 proteínas disponibles en NCBI, lo cual sugiere algún tipo de preferencia evolutiva para la conservación de este aminoácido (Figura 3.5.B). En una población segregante para la mutación E158K se observó que las plantas heterocigotas y homocigotas para esta mutación florecieron 5.1 ( $P = 0.0004$ ) y 9.1 ( $P = 2.15e^{-5}$ ) días más tarde respectivamente que el control sin la presencia de este cambio. Las plantas que no poseían este cambio de aminoácido florecieron sin diferencias significativas con la línea control TDF ( $P = 0.76$ , Figura 3.5.C). La segunda mutación analizada, corresponde con un cambio de nucleótido en el sitio dador de splicing del cuarto exón y fue encontrada en estado de homocigosis. La retención del intrón y su predicha traducción genera un codón de stop prematuro localizado a 8pm río abajo del sitio de splicing lo que elimina un 40% de la proteína. La presencia de este codón prematuro fue confirmada en los transcriptos secuenciados de este mutante (Materiales y Métodos). El ensayo de progenie de esta línea mostró un retraso en la floración de 146 días con respecto a la línea control TDF ( $P < 5.04e^{-14}$ ), y la floración ocurrió sin diferencias significativas con el control TDC ( $P = 0.910$ , Figura 3.5.D). Estos dos eventos independientes confirmaron que la copia duplicada de *VRN-A1* en la región proximal del cromosoma 5D es *VRN-D4*.

Las mutaciones en el fondo genético de los mutantes podrían influir en el efecto de las mutaciones caracterizadas, aunque se debe considerar que el carácter hexaploide de esta variedad actúa como ‘amortiguador’ de cualquier posible variación proveniente de mutaciones no deseadas. Además también el hecho de contar con dos mutaciones independientes que afectan el mismo gen hace que muy poco probablemente que los resultados observados sean resultados de mutaciones provenientes del fondo genético.

Tabla 3.1. Análisis de las mutaciones encontradas en *VRN-D4*.

		Mutaciones inducidas				
		exon 4	exon 6	exon 7		exon 8
<b>Línea</b>	TDF	Till-665	Till-526	Till-174	Till-481	Till278
<b>GenBank</b>	KR42242	KR11906	KR11906	KR11906	KR11906	KR11905
	4	3	2	1	0	9
<b>cambio nt.</b>	A367C	G10511A	G472A	G538A	G550A	G655A
<b>Cambio a.a.</b>	K123Q	Splice site	E158K	A180T	D184N	A219T
<b>PROVEAN<sup>a</sup></b>	-3.527	-	-3.292	-0.848	0.158	-0.306
<b>PolyPhen-2<sup>b</sup></b>	0.999	-	0.999	0.045	0.026	0.002
<b>SIFT<sup>c</sup></b>	No tolerado	-	No tolerado	tolerado	tolerado	tolerado
<b>PSSM<sup>d</sup></b>	1	-	-3	Fuera de K-box	Fuera de K-box	Fuera de K-box
<b>BLOSUM62</b>	1	-	1	-1	1	-1

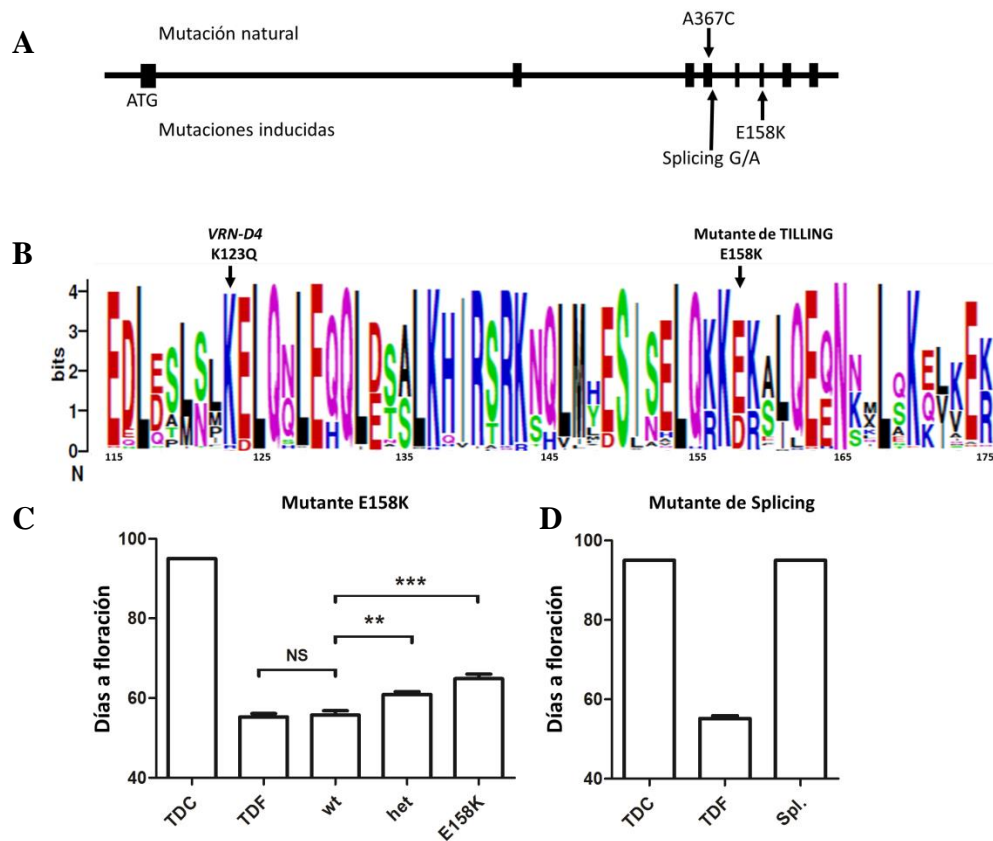
Las posiciones de cambio de nucleótido están calculadas desde el codón de inicio de la secuencia codificante de *VRN-D4*, con excepción del cambio en el sitio de splicing G10511A que es indicado relativo al codón de inicio a nivel de ADN en la secuencia genómica. Las posiciones de cambio de aminoácido son indicadas de acuerdo a la metionina inicial. Las letras antes y después del número indican el alelo salvaje y el mutante respectivamente.

<sup>a</sup> PROVEAN: los valores fueron calculados en [provean.jcvi.org](http://provean.jcvi.org). Un valor < -2.5 es predicho de afectar la función de la proteína.

<sup>b</sup> Phylophen-2: los valores fueron calculados usando el sitio web [genetics.bwh.harvard.edu/pph2/](http://genetics.bwh.harvard.edu/pph2/). Los valores cercanos a cero sugieren un efecto limitado en la función de la proteína, y los valores cercanos a 1 sugieren un efecto significativo en la estructura y función de la proteína.

<sup>c</sup> SIFT: fue calculado en [sift.jcvi.org](http://sift.jcvi.org) y los valores se resumen como cambios tolerados o no tolerados.

<sup>d</sup> Position-Specific Scoring Matrix (PSSM): este valor es basado en la conservación de los amino ácidos usando alineamientos múltiples de secuencias. Los valores debajo de cero indican cambios de aminoácidos que son predichos de tener un efecto significativo en la función de la proteína. El valor de PSSM puede ser calculado solamente para aquellas mutaciones dentro de dominios conservados, en este caso usando el dominio K-box (pfam01486) y la secuencia de proteína de *VRN-A1* (AAW73221.1) en NCBI ([www.ncbi.nlm.nih.gov/Class/Structure/pssm/pssm\\_viewer.cgi](http://www.ncbi.nlm.nih.gov/Class/Structure/pssm/pssm_viewer.cgi)).



**Figura 3.5. Análisis de las mutaciones inducidas generadas por TILLING en *VRN-D4*.**

(A) Estructura génica de *VRN-D4*, las flechas indican la localización del SNP A367C, y las mutaciones de TILLING (E158K y la mutación en el sitio de splicing) y la localización del SNP A367C. (B) Mediante un alineamiento múltiple de 268 secuencias que incluyeron los genes *API*, *AG*, *FUL*, *CAULIFLOWER*, *VRN1* y otras proteínas de la familia MAD-BOX se realizó la representación gráfica de la conservación de cada aminoácido. Se utilizó el sitio web de WebLogo ([weblogo.berkeley.edu](http://weblogo.berkeley.edu)) que a partir del alineamiento genera la representación gráfica donde en tamaño de cada letra representa un aminoácido y el tamaño es proporcional al su frecuencia en esas posición. La región mostrada cubre las posiciones del aminoácido 115 al 175 (en *VRN-D4*) dentro del dominio K-Box. Las flechas indican la posición donde se encuentran los cambios K123Q (natural de *VRN-D4*) y el producido mediante EMS E158K. (C) Caracterización de los mutantes de TILLING en tiempo a floración. Efecto de la mutación E158K en tiempo a floración. wt: homocigota para el alelo salvaje, het: heterocigota, E158K= homocigota para el alelo mutante. (D) Efecto de la mutación de splicing en tiempo a floración. Spl. = homocigota para la mutación de splicing. Las barras indican la media de al menos seis replicas biológicas. Las barras de error indican el error estándar de la media (SEM). NS = no significativo, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , y \*\*\* =  $P < 0.001$ ).

### 3.3.3 Mapa físico de la región de *VRN-D4*

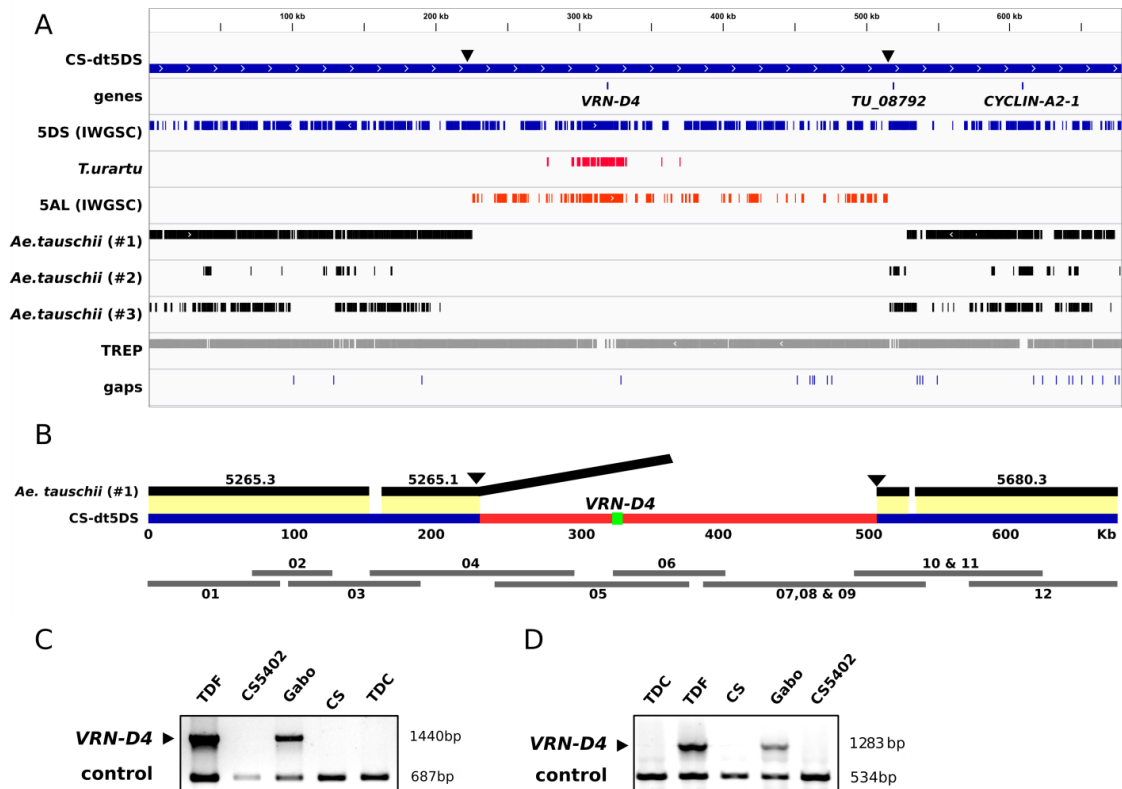
Mediante una búsqueda de BLAST en el sitio del borrador del genoma del trigo hexaploide Chinese Spring desarrollado por el consorcio internacional de secuenciación de trigo (IWGSC, <http://www.wheatgenome.org/Projects/IWGSC-Bread-Wheat-Projects/Physical-mapping/>) (Mayer et al. 2014) se encontró *VRN-D4* en la ubicación correspondiente al brazo corto del cromosoma 5D (SNP A367C presente). Aunque esta variedad, Chinese Spring, no posee *VRN-D4* la línea ditelisómica del cromosoma 5DS (CS-dt-5DS), utilizada para separar por cartometría de flujo y secuenciar este brazo cromosómico en el proyecto genoma, fue generada usando la variedad Gabo (Devos et al. 1999) la cual posee *VRN-D4*. Este resultado confirma la presencia de secuencias pertenecientes a Gabo en este stock genético como ya había sido sugerido con anterioridad (Pugsley 1972 y Goncharov 2003).

El mapa físico del brazo corto del cromosoma 5D fue utilizado para estudiar esta región cromosómica en detalle. Para esto se detectó el BACs que contenían *VRN-D4* mediante PCR, así como también los incluidos en el Minimum Tiling Path (MTP) de la biblioteca de BACs del cromosoma 5DS del Instituto de Botánica Experimental (República Checa) a cargo de producir estas para el IWGSC. Como resultado se encontró el BAC TaeCsp5DShA\_0038\_M14 el cual es parte de dos contigs mayores CTG87 y CTG61 (Tabla A.4). De estos se seleccionaron 12 BACs que rodean al BAC que posee *VRN-D4* y se secuenció usando una combinación de Illumina y 454 (Materiales y Métodos). Estos 12 BACs produjeron una secuencia de un total de ~680 Kb la cual fue analizada por similitud con otras secuencias de los borradores de proyectos genoma disponibles. Se utilizaron las bases de datos de secuencias de *T. urartu* (ancestro del genoma A de trigo hexaploide) (Ling et al. 2013), tres bases de datos de *Ae. tauschii* (ancestro del genoma D de trigo hexaploide) y las bases de datos de secuencia de los brazos cromosómicos 5DS y 5AL del IWGSC (ya que el brazo 5AL es la ubicación ancestral de *VRN-A1*). Mediante BLAT basados en una similitud mayor a 99% y longitud mayor a 500 pb se estudió la similitud de esta región con las distintas secuencias depositadas en las bases de datos descriptas más arriba y se detectaron tres regiones claramente definidas. La región central (~290 Kb) mostró alta similitud con el brazo cromosómico 5AL y *T. urartu*. Y las dos zonas que rodean la región central mostraron una alta similitud a las tres bases de datos del progenitor del genoma D

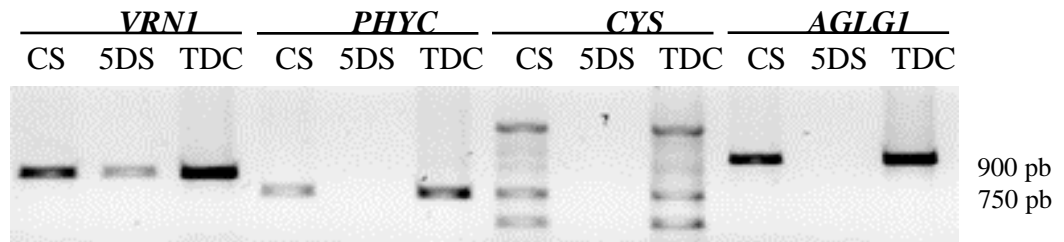
(Figura 3.6.A). En todos los casos las regiones mostraron una identidad única, ya sea con el genoma A o con el genoma D, sin detectarse regiones con similitud a ambos genomas. Este resultado indica que en este stock citogenético parece haber ocurrido una removilización génica del cromosoma 5A al cromosoma 5D, y por lo tanto haber dado origen a *VRN-D4*.

La anotación de la región secuenciada reveló la presencia de tres genes. *VRN-D4* fue el único gen identificado en la región con alta similitud al genoma A, y ninguno de los genes flanqueantes a *VRN1* en la ubicación original en el cromosoma 5AL (Yan et al. 2003) fue detectado mediante PCR. Los dos genes restantes encontrados están ubicados en la región con alta similitud al genoma D, río debajo de *VRN-D4* (*CYCLIN-A2-1* y un gen putativo TRIUR3\_08792). No se encontraron genes en la región ubicada río arriba de *VRN-D4*.





**Figura 3.6. Representación gráfica de la región cromosómica de *VRN-D4* en el brazo corto del cromosoma 5D.** (A) Ensamblado de la secuencia cromosómica CS-dt5DS (~680 kb) comparada con las bases de datos de secuencias de los brazos 5DS y 5AL (IWGSC), *T. urartu* y tres bases de datos de *Ae. tauschii*. Fueron anotados dos genes (*VRN-D4* y *Cyclin A2-1*= TRIUR3\_24995) y un gene hipotético TRIUR3\_08792. (B) Representación esquemática de la secuencia ensamblada: Amarillo= secuencia >99% idéntica a *Ae. tauschii*, Rojo= >99% idéntica al brazo 5AL; Negro = contigs de *Ae. tauschii* (<http://aegilops.wheat.ucdavis.edu/ATGSP/>), Rectángulos grises= BACs secuenciados, Amarillo= región >99% idéntica, Triángulos= inserción 5DS/5AL (C y D) Marcadores moleculares para la presencia de los sitios de inserción 5DS/5AL río arriba y río abajo de *VRN-D4* respectivamente. El producto de 1,440-pb indica la presencia de la 5DS/5AL río arriba de *VRN-D4* y la banda de 687-pb de tamaño corresponde a un control positivo de amplificación de la PCR (correspondiente al EST BJ315664). El producto de amplificación de 1,283-pb indica la presencia del sitio de inserción 5DS/5AL río abajo de *VRN-D4* y la banda de 534-pb corresponde a un control positivo de amplificación (EST BE606654).

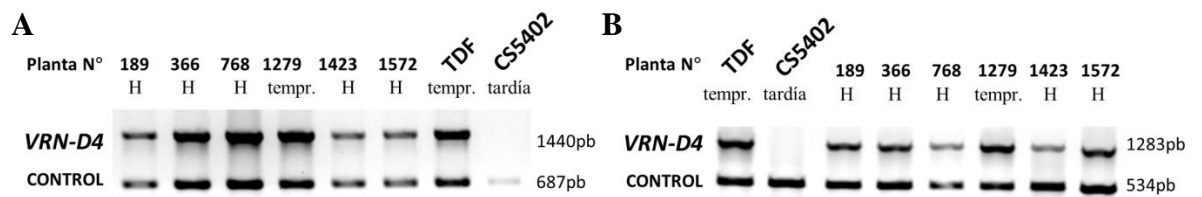


**Figura 3.7. Productos de amplificación de PCR de los genes flanqueantes a *VRN-A1* en ADN purificado del brazo cromosómico 5DS.** Se utilizó ADN del cromosoma 5DS purificado mediante cartometría de flujo (Instituto de Botánica Experimental, Olomouc, Republica Checa) para estudiar si los genes cercanos a *VRN-A1* en el cromosoma 5AL (Yan et al. 2003) se encuentran también presentes en la inserción en el cromosoma 5DS. CS: Chinese Spring, 5DS: ADN del brazo 5DS, TDC: Triple Dirk C. Usando el mismo ADN para todos los marcadores solamente *VRN-A1* fue amplificado, indicando que los genes *PHYC*, *CYS* y *AGLG1* no están presentes en el cromosoma 5DS. Las bandas múltiples en el gen *CYS* se deben a amplificaciones de las diferentes copias de este gen, ninguna presente en el ADN del brazo 5DS. Los genes *PHYC*, *CYS* y *AGLG1* tampoco fueron encontrados en la base de datos del cromosoma 5DS mediante BLAST, confirmado los resultados de PCR.

El sitio donde se produjo la inserción en el genoma D parece haber sufrido una alteración durante o después del evento de inserción. El contig de *Ae. tauschii* que presenta alta similitud con 104 Kb río arriba del sitio de inserción (contig 5265.1) contiene otros 200Kb de secuencia que no fueron encontrados río abajo del otro extremo de la inserción. Además esta región río abajo de la inserción presenta alta similitud con contigs diferentes en *Ae. tauschii* (Figura 3.6.B).

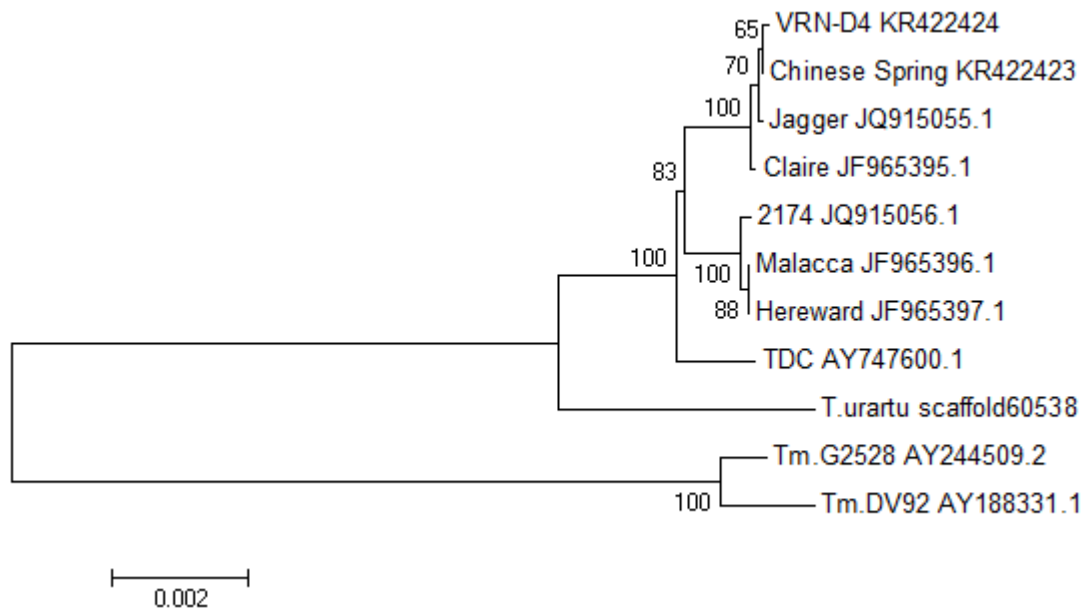
Las regiones de encuentro entre la alta similitud con el genoma A y el genoma D fueron utilizadas para diseñar marcadores de PCR para comprobar la existencia de estos eventos de inserción. El primer borde fue predicho río arriba de *VRN-D4*, ubicado a unos 88.3 Kb del codón de inicio. El segundo borde fue identificado entre 201 y 203 Kb río abajo de *VRN-D4*.

Los marcadores de PCR diseñados para cada borde mostraron estar perfectamente ligados con el SNP A367C diagnóstico de *VRN-D4* en las líneas portadoras del alelo primaveral de *VRN-D4* (Figura 3.8.C y D) y en las poblaciones descritas en la Tabla 3.2 y A.6. Estos resultados sugieren que *VRN-D4* detectado en diferentes líneas se originó por el mismo evento de inserción.



**Figura 3.8. Amplificación de PCR de los sitios de inserción en los recombinantes críticos de la población TDF x CS<sub>5402</sub>.** (A) Marcador para la inserción 5AL/5DS río arriba de *VRN-D4* (1,440-bp band). (B) Marcador para la inserción 5AL/5DS río abajo de *VRN-D4* (1,283-bp band). Nótese que todas las líneas portadoras del SNP A367C muestran amplificación de las bandas de 1,440pb y 1,283pb. H= segregante para el hábito de crecimiento (heterocigota *VRN-D4 vrn-D4*), tempr.= toda la progenie tuvo floración temprana (homocigota *Vrn-D4*), tardía= toda la progenie tuvo floración tardía (homocigota *vrn-D4*), control= como control positivo de la amplificación de PCR se incluyó un par de cebadores adicionales para producir amplificación en todas las líneas.

A continuación se estudiaron diferentes secuencias disponibles de *VRN-A1* así como las obtenidas en este estudio para detectar cual puede ser el posible ancestro en el cual la removiliación del cromosoma 5D al 5A tuvo lugar. Para esto se realizó un análisis de filogenia de siete diferentes alelos de *VRN-A1* incluyendo las secuencias de *T.monococcum* y *T.urartu* como referencia de los alelos ancestrales del genoma A. Como resultado de este análisis se encontró que el alelo más cercano a *VRN-D4* es el de *VRN-A1* presente en Chinese Spring, donde solo se encontró un SNP en 14,912pb secuenciados (SNP A367C). Las secuencias más cercanas a estas fueron las de las variedades Jagger y Claire, que mostraron uno y tres SNP adicionales respectivamente (Figura 3.9).



**Figura 3.9. Filogenia de *VRN-D4*.** Arbol filogenético basado en la región genómica de ADN de las secuencias de *VRN-A1* / *VRN-D4* (Neighbor-Joining utilizando 1000-pb río arriba del codon de inicio y 1000-pb río abajo del codón de stop de la transcripción) utilizando el programa MEGA6 (Tamura et al. 2013). Los números en los nodos de este árbol indican los valores de confianza basado en un 5000 submuestreos (% de árboles simulados con este nodo presente). Con excepción de *VRN-D4* todas las otras secuencias corresponden a *VRN-A1*. Los números indican los identificados correspondientes en GenBank, con excepción de la secuencia de *T. urartu*, en la cual se indica el número de contig en el ensamblado del proyecto genoma del mismo (plants.ensamble.org, GCA\_000347455.1).

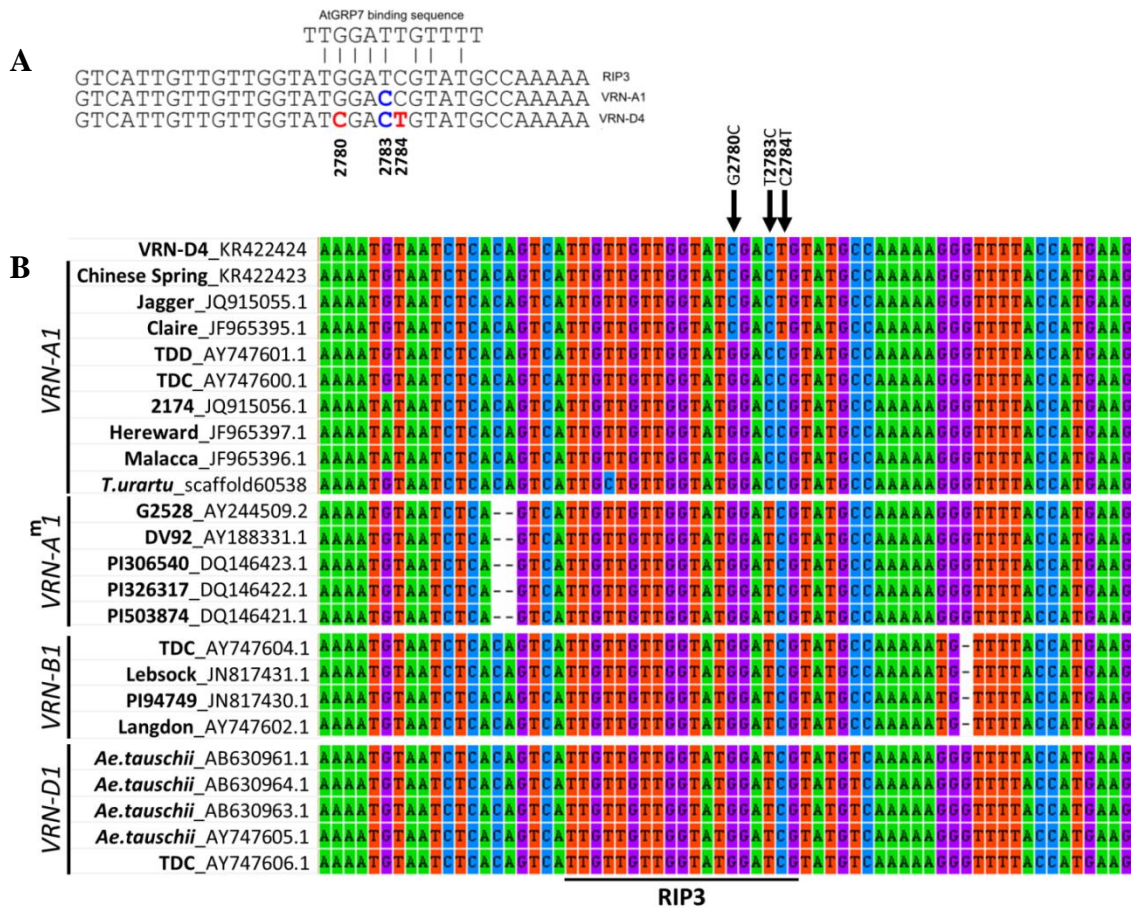
VRN-D4	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
Chinese_Spring	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
Jagger-JQ915055.1	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
Claire-JF965395.1	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
2174-JQ915056.1	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
TDC-Ay747600.1	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
Hereward-JF965397.1	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
Malacca-JF965396.1	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
Tm_G2528-Ay244509.2	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
T. urartu	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
T. turgidum-AAW73219	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
Tm_DV92-Ay188331.1	1	MGRGKVLKRIENKINRQVTFSKRRSGLLKAHEISVLCDAEVGLIIFSTKGKLYEFSTESCMDKILERYERYSYAEKVLVS	82
VRN-D4	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
Chinese_Spring	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
Jagger-JQ915055.1	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
Claire-JF965395.1	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
2174-JQ915056.1	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
TDC-Ay747600.1	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
Hereward-JF965397.1	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
Malacca-JF965396.1	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
Tm_G2528-Ay244509.2	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
T. urartu	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
T. turgidum-AAW73219	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
Tm_DV92-Ay188331.1	83	SESEIQGNWCHEYRKILAKVETIQKCQKHLMGEDLESNLKELQLEQQLLESSLKHRSRKNQMLMHESISELQKKERSLQEE	164
VRN-D4	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
Chinese_Spring	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
Jagger-JQ915055.1	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
Claire-JF965395.1	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
2174-JQ915056.1	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
TDC-Ay747600.1	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
Hereward-JF965397.1	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
Malacca-JF965396.1	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
Ay244509.2-Tm_G2528	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
T. urartu	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
T. turgidum-AAW73219	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245
Tm_DV92-Ay188331.1	165	NKVLQKELVEKQKAHAAQDDQTPQTSSSSSSFMLRDAPPAANTSIHPAATGERAEDAQVQAPPRTPGLPFWMVSHING*	245

**Figura 3.10. Análisis de la secuencia de proteína de VRN-D4.** Alineamiento múltiple de la secuencia de proteína predicha para VRN-D4 y las disponibles en bases de datos publicas VRN1. Resaltado en rojo se detalla el cambio de aminoácido característico de VRN-D4 K123Q. El dominio K-Box se indica en color gris. La secuencia de *T. urartu* fue obtenida del sitio web plants.ensembl.org (scaffold60538). La secuencia completa de VRN-A1 en la variedad Chinese Spring (GenBank KR422423) y VRN-D4 (GenBank KR422424) fueron obtenidas en este estudio.

### 3.3.4 Caracterización del alelo para hábito de crecimiento primaveral

Aunque VRN-D4 es expresado desde estadíos tempranos del desarrollo (primera hoja) en ausencia de vernalización, los resultados obtenidos del análisis de su secuencia a nivel de ADN no presentan ninguno de los polimorfismos en la secuencia promotora o el primer intrón previamente asociados con incrementos de la expresión génica en alelos de hábito de crecimiento primaveral (Yan et al. 2003 y 2004, Fu et al. 2005, Chu et al. 2011 y Zhang et al. 2012). Mediante un análisis comparativo que incluyó secuencias de VRN1 de los tres genomas de trigo hexaploide y de trigos diploides se observó que VRN-D4 posee tres SNPs en una región del primer intrón denominada RIP-3. Esta región ha sido recientemente identificada como crítica para la unión de un represor transcripcional de VRN1 llamado *TaGRP2* (Xiao et al. 2014). Se encontró la secuencia

canónica de RIP-3 en todas las secuencias correspondientes a *VRN-B1*, *VRN-D1* y en *VRN-A<sup>m1</sup>* (*T. monococcum*) (Figura 3.11). En contraste el genoma A presentó dos haplotipos diferentes. El primero corresponde a un cambio de T por C en la posición 2.783 desde el codón de inicio (denominado de aquí en más T2783C o RIP-3<sup>VRN-A1</sup>) y este fue también detectado cuando se estudió esta región mediante secuenciación en una colección de 33 *T. urartu*, 13 *T. turgidum* ssp. *dicoccoides*, 13 *T. turgidum* ssp. *dicoccum*, y 12 *T. aestivum* (Tabla A.5). El segundo haplotipo, encontrado en *VRN-D4* y en las secuencias de *VRN-A1* de las variedades Chinese Spring, Jagger y Claire (de aquí en más RIP-3<sup>VRN-D4</sup>), incluye el previamente descrito T2783C SNP pero también posee dos SNPs adicionales en las posiciones 2.780 (G2780C) y 2.784 (C2784T). Es interesante mencionar que variaciones en *VRN-A1* en las variedades Jagger y Claire han sido reportados previamente como ligados a una disminución en el requerimiento de vernalización (Díaz et al. 2012 y Li et al. 2013). Para comprobar si estos polimorfismos en la región RIP-3 afectan la unión de la proteína *TaGRP2* se utilizó un ensayo de interacción ARN-proteína (REMSA, descrito en la sección de materiales y métodos).



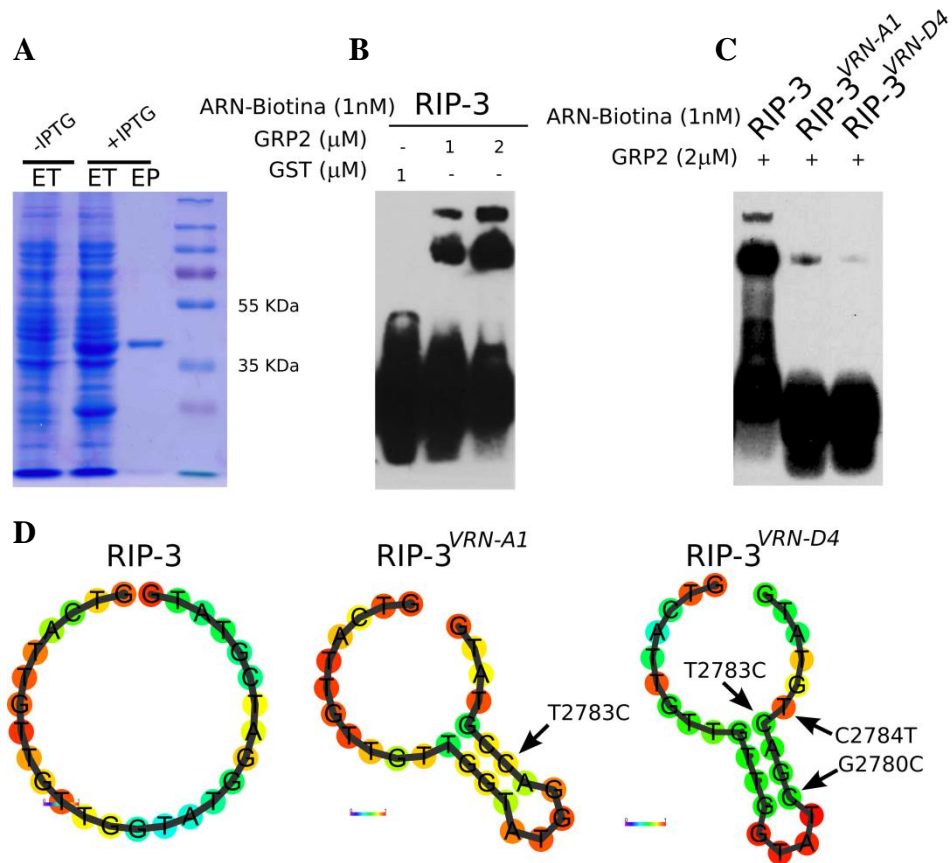
**Figura 3.11. Detalle de la región correspondiente a RIP-3 en el primer intrón de *VRN-1* y *VRN-D4*.**

(A) El sitio de unión de GRP7 en Arabidopsis (Streitner et al. 2008) comparte 8 nucleótidos con la secuencia de RIP-3 donde se une TaGRP2. (B) La secuencia canónica de RIP-3 fue encontrada en *VRN-B1*, *VRN-D1* y *VRN-A<sup>m</sup>1* (*T. monococcum*), mientras que *VRN-A1*, *T. urartu* y todas las líneas de trigo tetraploide y hexaploide secuenciadas en la Tabla A.5 comparten el cambio de nucleótido T2783C. El haplotipo de RIP-3 con 3 SNPs fue encontrado en *VRN-D4* y en las secuencias de *VRN-A1* de las variedades Chines Spring, Jagger y Claire, el cual fue asociado con un débil requerimiento de vernalización. Las flechas en RIP-3 indican la posición de los polimorfismos 2780, 2783 y 2784 (contando desde el ATG de *VRN-D4*). Los nombres de las secuencias incluyen los números de acceso de GenBank.

En primer lugar se analizó la unión de *TaGRP* a un fragmento de ARN de 27-nt idéntico al previamente descrito para RIP-3 por Xiao y colegas (2014) (Figura 3.12.B) dicho fragmento corresponde a la secuencia ‘canónica’ para este sitio, e igual a la que se encontró en este estudio de las secuencias de *VRN1* en los genomas B, D y en *T.monococcum*. Esta fue comparada con sondas de ARN de 27-nt conteniendo los SNPs presentes en RIP-3<sup>VRN-A1</sup> y RIP-3<sup>VRN-D4</sup>. Como resultado se observó que *TaGRP2* se unió a RIP-3 de forma dosis-dependiente (Figura 3.12.B). La intensidad de la unión decreció



cuando se utilizó la sonda con la presencia del SNP T2783C presente en  $RIP-3^{VRN-A1}$  y fue casi totalmente eliminada en presencia de los tres SNPs característicos de  $RIP-3^{VRN-D4}$  (Figura 3.12.C).



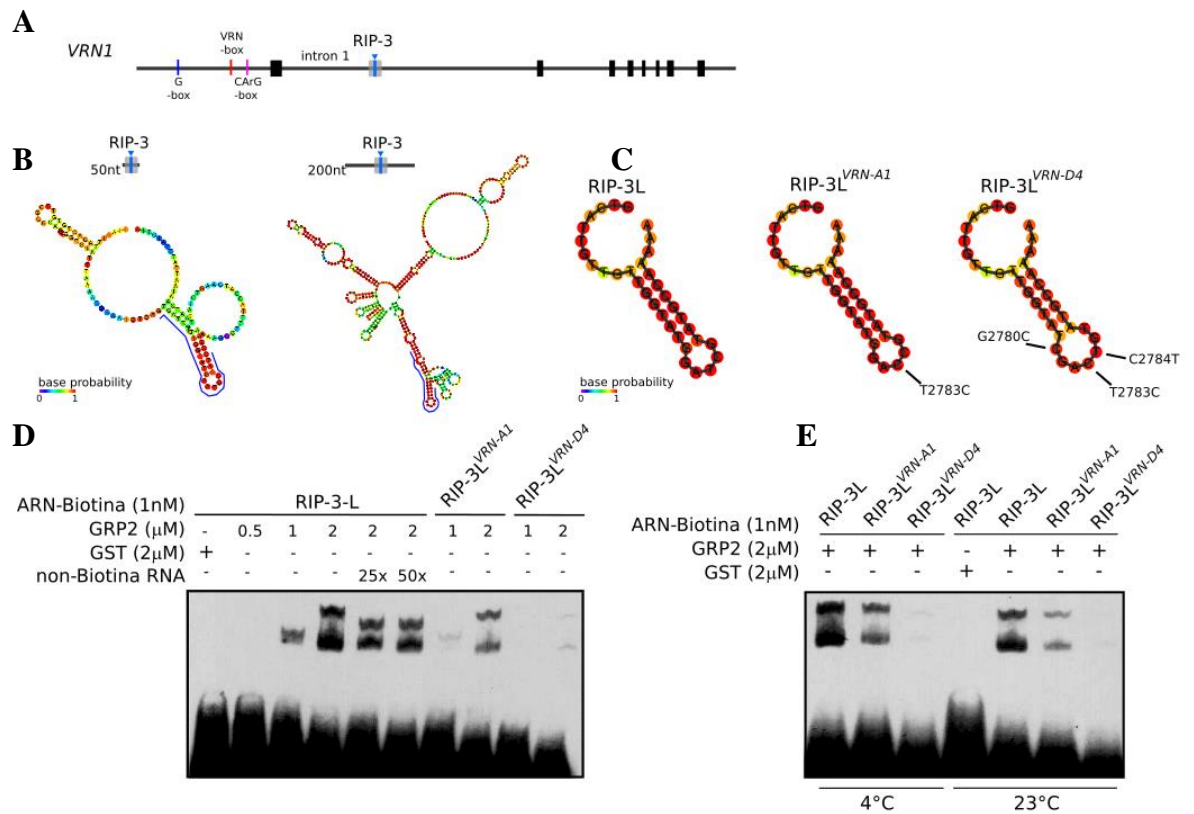
**Figura 3.12. Unión de TaGRP2 a la región RIP-3 en el primer intrón de *VRN1/VRN-D4*.** (A) Expresión de GST-TaGRP2 en *E. coli*, ET: Extracto total, EP: elución de proteína. (B) Ensayo de movilidad electroforética de ARN (REMSA) mostrando la interacción entre TaGRP2 y la sonda de 27-nt de ARN de RIP-3, usando la misma secuencia publicada previamente (Xiao et al 2014). Fueron analizadas dos concentraciones de TaGRP2 (1= 1 $\mu$ M y 2= 2 $\mu$ M) (C) REMSA mostrando las interacciones entre TaGRP2 con las tres diferentes sondas de ARN de 27-nt: RIP-3 presentó una interacción fuerte, se observó una reducción en la interacción con la sonda RIP-3<sup>VRN-A1</sup> (SNP T2783C), y se detectó una interacción muy débil con la sonda RIP-3<sup>VRN-D4</sup> (SNPs G2780C, T2783C y C2784T, también presente en *VRN-A1* en Jagger, Claire y Chinese Spring). (D) Estructuras secundarias predichas para las sondas de 27-nt utilizadas mediante el programa RNA fold (Gruber et al. 2008). Las flechas indican la posición de los tres SNPs desde el codón de inicio de la traducción en *VRN-D4* (2780, 2783 y 2784).

A continuación se estudió la estructura secundaria predicha para la secuencia de ARN de 27-nt y se detectó que la estructura es diferente entre las sondas que poseen los SNPs y la secuencia canónica de RIP-3. Estas estructuras predichas fueron diferentes de las



estructuras observadas cuando se utilizó una secuencia flanqueante a RIP-3 de mayor longitud (Figura 3.12.D), sugiriendo que podrían no representar la estructura real adoptada por el ARNm de *VRN1* al momento de interactuar con *TaGRP2* y que los SNPs introducidos podrían afectar la estructura secundaria del ARN. Para descartar esta posibilidad se utilizaron sondas de mayor longitud (34nt, RIP-3L) que adoptan estructuras secundarias similares a las observadas en fragmentos de mayor longitud (Figura 3.13.B-C). Con estas sondas de 34-nt se repitieron los ensayos de interacción con *TaGRP2* (Figura 3.13.D), encontrando nuevamente que la sonda correspondiente a la secuencia canónica de RIP-3L fue la que mostró mayor unión a *TaGRP2*. RIP-3L<sup>VRN-A1</sup> mostró una ligera disminución de la unión, y RIP-3L<sup>VRN-D4</sup> eliminó la unión a *TaGRP2* casi completamente. Resultados similares se encontraron cuando este ensayo fue repetido a una temperatura de unión de 4°C (Figura 3.13.E).

En resumen, las sondas con diferentes longitud y las diferentes temperaturas de unión mostraron el mismo patrón de unión con *TaGRP2*, en forma decreciente desde RIP-3, RIP-3<sup>VRN-A1</sup>, a RIP-3<sup>VRN-D4</sup>.



**Figura 3.13. Interacción de TaGRP2 con las sondas de ARN del sitio RIP-3.** (A) Sitio de unión de GRP2 en el primer intrón de *VRN1*. (B) Estructura secundaria predicha para la secuencia de ARN incluyendo 50 nt o 200 nt para cada lado del sitio RIP-3. (C) Estructura secundaria predicha de las tres sondas de ARN de 34-nt utilizadas (-L). RIP-3L: secuencia canónica, RIP-3<sup>VRN-A1-L</sup>: SNP T2783C, RIP-3<sup>VRN-D4-L</sup>: SNPs G2780C, T2783C y C2784T. Las flechas indican las posiciones de los SNPs estudiados en la estructura predicha y los colores representan la probabilidad del apareamiento de las bases. (D) Ensayo de movilidad electroforética del ARN (REMSA) mostrando la interacción de TaGRP2 con las tres sondas de ARN. Las interacciones se estudiaron utilizando dos concentraciones de proteína (1= 1 μM y 2= 2 μM). Una sonda RIP-3L sin marcar fue utilizada como competidor. (E) Las mismas sondas utilizadas en D fueron estudiadas a dos diferentes temperaturas de unión (calles 1-3= 4 °C y calles 4-7= 23 °C).

### 3.3.5 Distribución geográfica de *VRN-D4*

Los polimorfismos en la región reguladora RIP-3 presentes en *VRN-D4* fueron estudiados en la población de mapeo fino TDFxCS<sub>5402</sub> detallada en el capítulo 1, y fueron encontrados perfectamente ligados con el SNP A367C SNP de la región codificante de *VRN-D4* y también con los dos marcadores para la inserción 5DS/5AL. Los tres SNPs de la región RIP-3 también estuvieron ligados con el SNP A367C SNP en una selección de 26 líneas de *T. aestivum* que han sido previamente sugeridas como

portadoras de *VRN-D4* basado en cruzamientos y análisis de segregación genética (Goncharov 1999 e Iwaki et al. 2000) (Tabla 3.2). De estas 26 líneas, 22 presentaron los cuatro polimorfismos diagnósticos de *VRN-D4*. Las cuatro líneas restantes no presentaron ninguno de estos polimorfismos, sugiriendo una incorrecta asignación de *VRN-D4*. El estudio del número de copias de *VRN-A1* en estas líneas mostró que estas cuatro no tienen copias extras de *VRN-A1* fuera del sitio ancestral de *VRN1* en el brazo largo del cromosoma 5A, confirmando la ausencia de *VRN-D4*. Las líneas IL193 y CL035 mostraron un número de copias de *VRN-A1* igual a uno, mientras que las líneas IL346 y CL030 mostraron un número de copias igual a dos, pero también la presencia de un SNP diagnóstico para una duplicación en tándem de *VRN-A1* en el cromosoma 5AL (SNP C349T, Tabla S4) (Díaz et al. 2012).

El sitio de colección de estas líneas que presentan *VRN-D4* (Tabla 3.2) se solapa parcialmente con el sitio de distribución de *T. aestivum* ssp. *sphaerococcum* (región de Punjab, actualmente Noreste Pakistán y Noroeste de la India, Figura 3.14), una subespecie de *T. aestivum* que ha sido propuesta como posible fuente de *VRN-D4* (Goncharov 2003). La caracterización de una colección de 33 entradas de *T. aestivum* ssp. *sphaerococcum* para la presencia del SNP A367C y los dos marcadores para la inserción 5DS/5AL confirmaron la presencia de *VRN-D4* en 31 de las 33 entradas (Tabla 3.3). Las únicas dos entradas clasificadas como *T. aestivum* ssp. *sphaerococcum* donde no se encontró ninguno de los marcadores para *VRN-D4* fueron colectadas al Noreste de China (Citr 8610 y Citr 10911), fuera del área de distribución central de esta subespecie. Estas dos líneas también difieren de las otras entradas de *T. aestivum* ssp. *sphaerococcum* por la presencia del alelo primaveral de *Vrn-D1* (Tabla 3.3), el cual

**Tabla 3.2. Distribución geográfica de las entradas postuladas como portadoras de *VRN-D4* en**

**estudios previos.** La presencia de los productos de los marcadores para los bordes de la inserción 5AL/5DS, y la presencia de la secuencia de RIP-3<sup>VRN-D4</sup> es indicada como “+” y la ausencia como “-”. La detección del número de copias de *VRN-A1* (CNV) es descrita en la sección de materiales y métodos. La duplicación en tándem de *VRNAI* fue estudiada mediante la presencia/ausencia de un doble pico C/T en la posición 349 de la región codificante de *VRN-A1* como se ha descrito con anterioridad (Díaz et al. 2013). TDF, GABO, CS<sub>5402</sub> y TDC fueron incluidos como controles. Los nombres de las líneas corresponden a los publicados en estudios previos (Iwaki et al. 2000 y 2001). Los nombres comenzando con la letra ‘N’ corresponden a entradas del Instituto de Citología y Genética (Novosibirsk, Rusia).

Nombre	Borde superior	Borde inferior	<i>VRN-D4</i> SNP A367C	RIP-3 <i>VRN-D4</i>	CNV	Duplicación <i>VRN-A1</i>	Locación
TDF	+	+	A/C	+	2	ND	-
GABO (N6827)	+	+	A/C	+	3	ND	Australia
CS <sub>5402</sub>	-	-	A	(+) <sup>1</sup>	1	ND	-
TDC	-	-	A	-	1	ND	-
CL035	-	-	A	(+) <sup>1</sup>	1	ND	China (south east)
IL193	-	-	A	(+) <sup>1</sup>	1	ND	Ethiopia
CL030	-	-	A	-	2	presente	China (south east)
IL346	-	-	A	-	2	presente	Afghanistan (west)
IL165	+	+	A/C	+	3	ND	Egypt
IL430	+	+	A/C	+	2	ND	Pakistan (north)
SS23	+	+	A/C	+	3	ND	Australia
IL027	+	+	A/C	+	2	ND	Afghanistan
IL047	+	+	A/C	+	2	ND	Turkey
IL114	+	+	A/C	+	2	ND	Nepal
IL154	+	+	A/C	+	4	ND	Pakistan (south)
IL163	+	+	A/C	+	3	ND	Egypt
IL213	+	+	A/C	+	2	ND	Nepal
IL216	+	+	A/C	+	2	ND	Nepal
GR007	+	+	A/C	+	2	ND	India
GR014	+	+	A/C	+	2	ND	India
GR015	+	+	A/C	+	2	ND	India
GR023	+	+	A/C	+	2	ND	India
IL164	+	+	A/C	+	3	ND	Egypt
IL175	+	+	A/C	+	3	presente	Bhutan
IL176	+	+	A/C	+	3	presente	Bhutan
IL008	+	+	A/C	+	3	presente	Nepal
IL431	+	+	A/C	+	4	presente	Pakistan (north)
GR001	+	+	A/C	+	3	presente	India
N7451	+	+	A/C	+	2	ND	
N6721	+	+	A/C	+	3	presente	

<sup>1</sup> Estas líneas poseen el alelo RIP-3<sup>VRN-D4</sup> en *VRN-A1*. ND= No detectado.

posee una delección del primer intrón asociada con hábito de crecimiento primaveral altamente frecuente en estas zonas de cultivo de trigo en China (Zhang et al. 2008). Para estudiar si la ausencia de *VRN-D4* en estas dos líneas es el resultado de mezcla con otras subespecies de *T. aestivum*, esta población fue comparada con cinco diferentes subespecies de *T. aestivum* (Tabla A6) usando un chip de 90 mil marcadores SNP (90K iSelect SNP chip (Wang et al. 2014)). Los resultados fueron analizados mediante estudios de similitud y estructura de la población. Como resultado se observó que estas cinco subespecies se agruparon en seis subgrupos, incluyendo, dos grupos para *T. aestivum* ssp. *sphaerococcum* y un grupo que combinó *T. aestivum* ssp. *aestivum* con entradas de *T. aestivum* ssp. *compactum* (Figura 3.17). En general hubo una buena correlación entre la caracterización previa y los agrupamientos de las subespecies. Un número menor de entradas muestra una probabilidad más alta de pertenecer a un grupo diferente al asignado previamente (Figura 3.17).

Por ejemplo, tres entradas caracterizadas como *spelta*, muestran una probabilidad más alta de pertenecer al grupo de *sphaerococum* o *compactum* que a su propio grupo. Así mismo, algunas entradas de *sphaerococcum* comparten una mayor probabilidad de pertenecer al grupo de *aestivum*. Las entradas colectadas en China CItr 8610 y CItr 10911 mostraron una gran evidencia de mezcla con otras subespecies lo cual puede explicar la ausencia de *VRN-D4* en estas dos líneas (asteriscos Figura 3.16 y Figura 3.17.C). En términos generales, la presencia de diferentes grupos comprendiendo las diferentes subespecies muestra un limitado flujo genético entre *T. sphaerococcum* y el resto de las subespecies.



Resultados similares se observaron cuando se calculó la matriz de correlación, con un grupo claramente identificable formado por *sphaerococcum*, otros dos grupos bien definidos para las subespecies *macha* y *spelta*, y un tercer grupo con gran similitud formado entre las subespecies *aestivum* y *compactum*. Cuando se calculó el índice de diferenciación ( $F_{st}$ ), también se obtuvieron resultados similares, donde *aestivum* y *compactum* tienen los índices de diferenciación más bajos, lo que implica alta similitud entre estas subespecies, y *sphaerococcum* mostró tener el mayor índice de diferenciación en comparación con las otras cuatro subespecies (Figura 3.17.D), lo que indica que esta subespecie es la más diversa entre las incluidas en este estudio. El hecho de que *VRN-D4* esté presente en todas las entradas de *T. aestivum* ssp. *sphaerococcum* de la región de Punjab, sugiere que probablemente este haya contribuido a la adaptación local de esta subespecie y haya sido favorecido mediante selección positiva. Para estudiar esta hipótesis caracterizamos la diversidad genética promedio (Weir 1996) de cada cromosoma y ciertas regiones cromosómicas en las diferentes subespecies de *T. aestivum* usando 14.236 marcadores para los cuales conocemos su ubicación cromosómica. Para reducir sesgos en los cálculos de la diversidad, los valores de diversidad por cromosoma fueron estandarizados por la diversidad del correspondiente genoma y subespecie (Tabla 3.4). Los valores obtenidos referidos como diversidad genética estandarizada y son mostrados en la Figura 3.15.

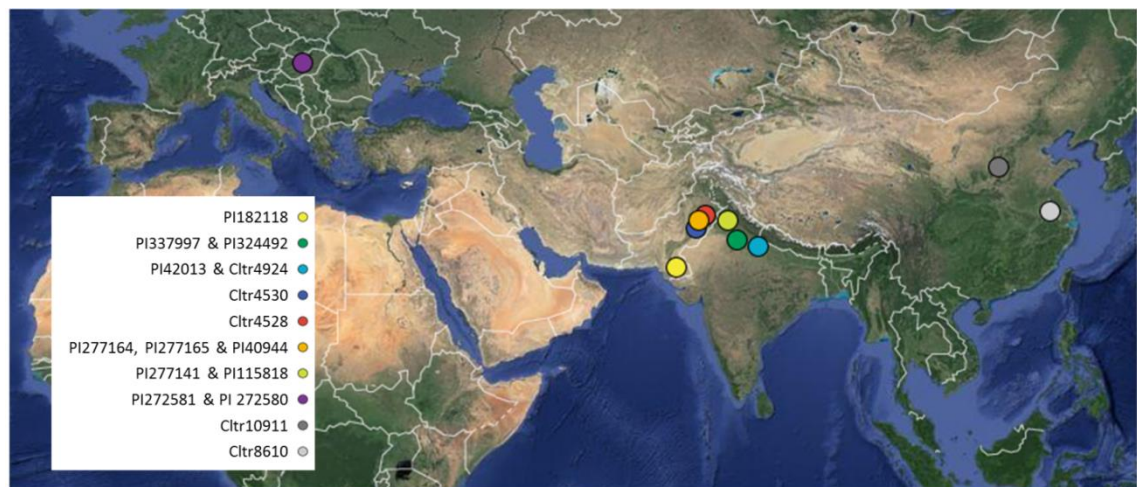
Se encontró que el cromosoma 5D de *T. aestivum* ssp. *sphaerococcum* presenta una diversidad genética de casi la mitad de la diversidad observada para las otras cuatro subespecies de *T. aestivum*. En ninguno de los otros 20 cromosomas se observó una reducción semejante entre subespecies (Figura 3.15). Para estudiar como distintas regiones del cromosoma 5D aportan a la reducción en diversidad genética observada se dividió el cromosoma en dos regiones, una proximal (1/3 de la distancia genética que incluye *VRN-D4*) y una región distal (que incluye 2/3 de la distancia genética). El mismo análisis se realizó para los cromosomas 5A y 5B (Figura 3.17.E). Cuando fue comparado con el promedio de la diversidad genética en las otras cuatro subespecies, el cromosoma 5D de *T. aestivum* ssp. *sphaerococcum* mostró una reducción de 1.8-veces en la diversidad genética en la región distal y una reducción de 4.7-veces de la diversidad genética en la región proximal a *VRN-D4* (Fig. 5E). Este patrón no fue observado en los cromosomas 5A o 5B. (Figura 3.17.E). En resumen, estos resultados indican que existe una reducción en la diversidad genética del cromosoma 5D de *T.*

*aestivum* ssp. *sphaerococcum* y que la región proximal de este cromosoma, que incluye a *VRN-D4*, es responsable por la mayor parte de esta reducción en diversidad genética.

**Tabla 3.4. Diversidad genética por genoma de las cinco subespecies de *T. aestivum*.** Se utilizó un total de 14,263 SNPs para los cuales se conoce su ubicación cromosómica seleccionados del conjunto inicial de marcadores de 16,371 SNPs identificados en el chip 90K iSelect Illumina (Wang et al. 2014) para estudiar la diversidad genética de cada genoma.

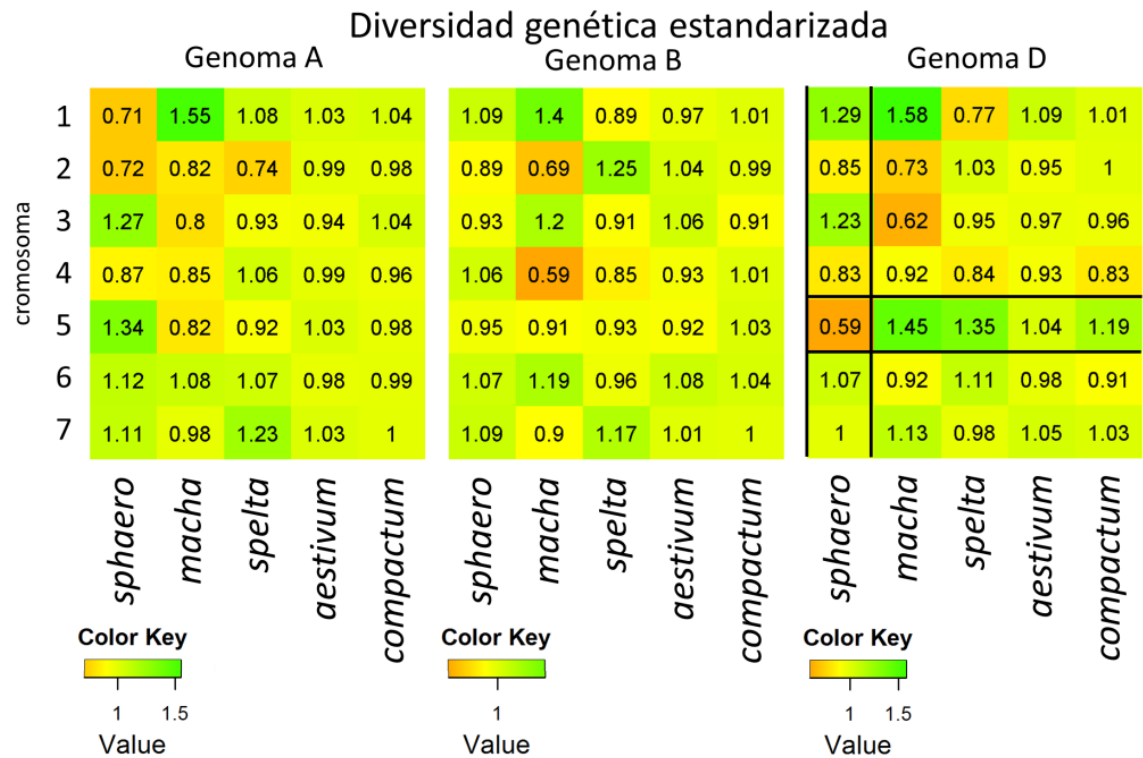
	No. de SNPs	<i>sphaerococcum</i>	<i>macha</i>	<i>spelta</i>	<i>aestivum</i> *	<i>compactum</i> *
Genoma A	5675	0.202	0.216	0.226	0.388	0.390
Genoma B	7344	0.243	0.189	0.215	0.388	0.382
Genoma D	1217	0.269	0.156	0.283	0.385	0.354

La diversidad genética se calculó como se detalla en la sección de Materiales y Métodos de este capítulo  
 \*Estos SNPs fueron desarrollados para *T. aestivum* ssp. *aestivum*, por lo que la alta diversidad genética observada en esta subespecie y en la subespecie *T. aestivum* ssp. *compactum*, puede ser influenciada por sesgos en el desarrollo de esta plataforma. Por lo que estos datos de diversidad son solamente utilizados para ajustar la diversidad genética de distintas regiones cromosómicas en cada subespecie y de esta forma contar con valores estandarizados de diversidad genética en futuros análisis.

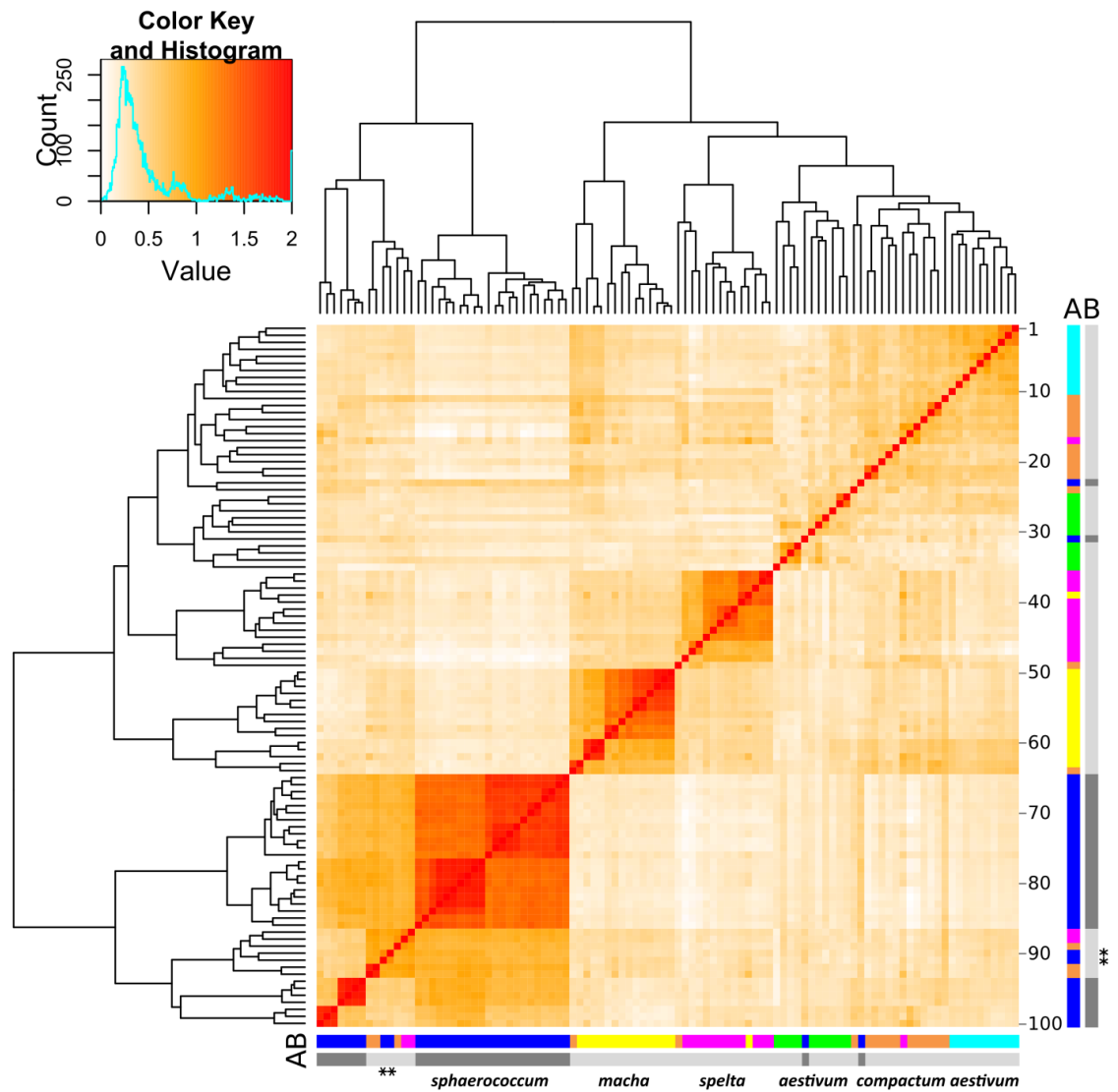


**Figura 3.14. Sitio de colección geográfica de las entradas de *T. aestivum* ssp. *sphaerococcum* utilizadas en este estudio.** Los colores representan cada localización geográfica de acuerdo a las coordenadas disponibles en GRIN ([www.ars-grin.gov/npgs/searchgrin.html](http://www.ars-grin.gov/npgs/searchgrin.html)). Los puntos fueron posicionados de acuerdo a las coordenadas geográficas utilizando Google Maps ([maps.google.com](http://maps.google.com)). Todas las entradas poseen *VRN-D4* con excepción de las entradas Cltr 8610 y Cltr 10911, ambas colectadas en China del este (gris).

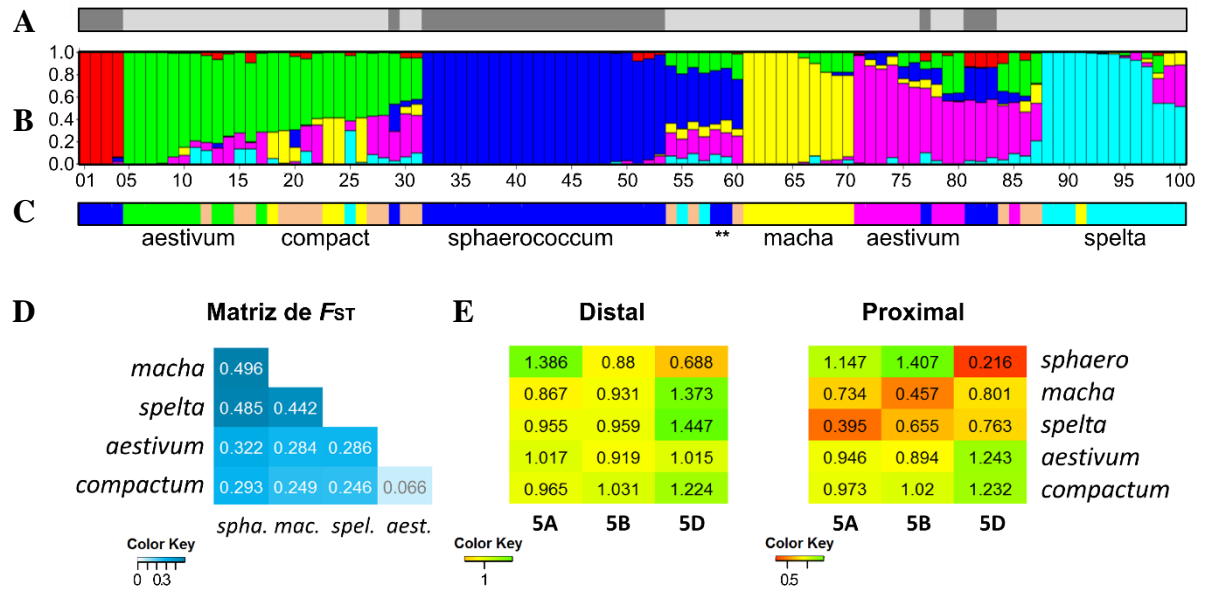




**Figura 3.15. Diversidad genética por subespecie y cromosoma.** La diversidad genética fue calculada para cada locus (14,236 SNP) y cada subespecie usando la fórmula del Contenido de Información Polimórfica (PIC) descrita en la sección de materiales y métodos. Estos valores fueron promediados para cada cromosoma, y para reducir los efectos de sesgos en el diseño de esta plataforma, se dividió este promedio por la diversidad genética promedio del genoma y la subespecie correspondiente. Nótese que el cromosoma 5D de *T. aestivum* ssp. *sphaerococcum* (*sphaero.*) es uno de los valores más bajos de diversidad genética, y este cromosoma muestra una diversidad genética más alta en las otras cuatro subespecies.



**Figura 3.16. Representación de la matriz de correlación de las 100 entradas de *T. aestivum* (5 subespecies) utilizadas en este estudio.** Este análisis fue basado en 16.371 SNPs. Rojo= alta similitud. Los números a la derecha indican los identificadores mostrados en la Tabla A.6 y la Figura 3.17. (A) En la barra exterior se indica la presencia o ausencia de *VRN-D4* mediante color gris oscuro y claro respectivamente. (B) En la barra de color se indican las entradas de: *T. aestivum* ssp. *aestivum* hábito de crecimiento primaveral en verde, y con hábito de crecimiento invernal en celeste, *T. aestivum* ssp. *compactum* en color naranja, *T. aestivum* ssp. *sphaerococcum* in en azul oscuro, *T. aestivum* ssp. *macha* en amarillo, y *T. aestivum* ssp. *spelta* con color rosado. Los asteriscos indican las dos entradas descritas como *T. aestivum* ssp. *sphaerococcum* colectadas en China y que no poseen *VRN-D4* (Citr 8610 y Citr 10911). La clasificación taxonómica se corresponde con la información disponible en GRIN ([www.ars-grin.gov/npgs](http://www.ars-grin.gov/npgs)).



**Figura 3.17. Relaciones genéticas entre las subespecies de *T. aestivum* estudiadas.** (A) Gris oscuro indica la presencia y gris claro la ausencia de *VRN-D4* basado en los marcadores desarrollados en este estudio. (B) Estructura de la población basada en 16371 marcadores SNP. Cada barra representa un genotipo (100 líneas en total). (C) Clasificación taxonómica de acuerdo a GRIN ([www.ars-grin.gov/npgs](http://www.ars-grin.gov/npgs)). Verde: *aestivum* (hábito de crecimiento primaveral), naranja: *compactum*, azul: *sphaerococcum*, amarillo: *macha*, rosa: *aestivum* (hábito de crecimiento invernal), cian: *spelta*. Los asteriscos indican las dos entradas de clasificadas como ssp. *sphaerococcum* colectadas en China donde *VRN-D4* no está presente. Los identificadores de las entradas están descriptos en la Tabla A.6. (D) Nivel de diferenciación genética entre todas las combinaciones de pares de subespecies determinada mediante el cálculo de índice de diferenciación ( $F_{ST}$ ). Todas las comparaciones son significativas con un  $P < 0.001$ . (E). Valores de diversidad genética estandarizada para la región distal y proximal de los cromosoma del grupo 5. Nótese la alta reducción de la diversidad en la región proximal del cromosomas 5D en *T. aestivum* ssp. *sphaerococcum* relativa a las otras subespecies.

## 3.4 DISCUSIÓN

### 3.4.1 Un genoma dinámico

En este trabajo se mostró que *VRN-D4* se originó mediante la removilización de una gran región genética del brazo largo del cromosoma 5A, la cual incluyó a *VRN-A1*, al brazo corto del cromosoma 5D. La removilización de genes y fragmentos génicos a ubicaciones no colineares es altamente frecuente en los grandes genomas de los pastos de clima templado, los cuales se originaron de amplificaciones masivas de elementos repetitivos (38, 46-48). Algunos transposones de ADN en trigo (como los CACTA) son capaces de removilizar genes y fragmentos génicos a regiones no colineares (Wicker et al. 2003), como ha sido también reportado en arroz para los ‘Pack-MULEs’ (Jiang et al. 2004) y también en maíz para los ‘Helitrons’ (Brunner et al. 2005)). Sin embargo, este mecanismo está asociado con fragmentos de ADN relativamente pequeños, menores a 3 Kpb (Wicker et al. 2010), y es poco probable que sea el mecanismo mediante el cual se produjo la inserción del fragmento de 290K pb que originó a *VRN-D4*.

Existe un mecanismo alternativo propuesto para explicar removilizaciones de ADN que involucran la inserción de fragmentos de ADN foráneos usados como molde para reparar roturas de doble cadena del ADN generadas por la inserción de elementos repetitivos (Wicker et al. 2010). Una huella particular de este mecanismo es la presencia de inserción de retrotransposones en los bordes del sitio aceptor (Wicker et al. 2010). Este mecanismo está asociado con un amplio rango de tamaños de insertos, y por lo tanto podría explicar la inserción del fragmento de 290 Kpb del cromosoma 5AL al 5DS. Desafortunadamente, uno de los bordes originales río debajo de la inserción ha sido alterado, limitando este tipo de análisis para determinar el mecanismo que origino esta inserción.

Más allá del mecanismo mediante el cual ocurrió este evento, la inserción de un segmento del cromosoma 5AL en el 5DS provee un ejemplo bien documentado de la capacidad de re movilización de grandes regiones cromosómicas a diferentes localizaciones cromosómicas en tiempos históricos.

Adicionalmente nuestros resultados demuestran que las secuencias producidas en el proyecto genoma de trigo deben ser interpretadas cuidadosamente. Debido a que la

estrategia utilizada para reducir la complejidad de secuenciación de un genoma de trigo (125 veces más grande que el de *Arabidopsis*) es preciso considerar que algunas de las secuencias catalogadas como contaminaciones pueden representar rearrreglos genéticos reales con relevancia biológica. Nuestros resultados demuestran que algunas de las secuencias depositadas como pertenecientes a la variedad Chinese Spring (Mayer et al. 2014) realmente corresponden a una variedad diferente. Así también debe ser considerada la presencia, que ha sido reportada con anterioridad (Devos et al. 1999), de otros fondos genéticos diferentes a Chinese Spring en algunos de las líneas aneuploides utilizadas en el proyecto genoma.

### 3.4.2 El origen de *VRN-D4*

En previos estudios genéticos se había sugerido que algunos individuos pertenecientes a la subespecie *T. aestivum* ssp. *sphaerococcum* eran portadores de *VRN-D4*, catalogando a este mismo como la posible fuente de este alelo primaveral (Stelmakh et al. 1996, Goncharov y Shitova 1999 y Goncharov 2003). Esta subespecie, descrita por primera vez hace más de noventa años (Percival 1921), ha sido encontrada distribuida principalmente en la región de Punjab en el sur de Asia, esta región comprende la región del este de Pakistán y el noroeste de la India. Los granos característicos de *T. aestivum* ssp. *sphaerococcum* con forma semiesférica son abundantes en los sitios arqueológicos de lo que se conoce como el Valle de la India, que fue la el sitio principal de la civilización que lleva el mismo nombre, en un periodo de hace aproximadamente entre 4600 y 3900 años (Weber 1999).

La identificación y caracterización de *VRN-D4* nos permitió estudiar su distribución en una colección de germoplasma más abundante en la cual *VRN-D4* fue encontrado presente en 31 de las 33 entradas de *T. aestivum* ssp. *sphaerococcum*, pero ausente en todas las entradas de las otras subespecies de *T. aestivum* estudiadas en este trabajo. Estos resultados soportan la hipótesis de que *VRN-D4* se originó en *T. aestivum* ssp. *sphaerococcum*.

### 3.4.3 El efecto de *VRN-D4* en la diversidad genética de *T. aestivum* ssp. *sphaerococcum*

La alta frecuencia con la que *VRN-D4* está presente en la colección de *T. aestivum* ssp. *sphaerococcum* y la ausencia de otros *loci* para hábito de crecimiento primaveral sugieren que *VRN-D4* se originó temprano en la evolución de esta subespecie o que este gen fue sujeto de una intensa selección. Para investigar estos distintos escenarios, se comparó la distribución genética en cada cromosoma y en diferentes regiones cromosómicas de las distintas subespecies de *T. aestivum*. La reducción de entre 4 a 5 veces de la diversidad genética en la región proximal del cromosoma 5D de *T. aestivum* ssp. *sphaerococcum* relativa a las otras cuatro subespecies sugiere que la región de *VRN-D4* estuvo sujeta a una intensa selección positiva (Figura 3.16 y Figura 3.17). En el caso de contar con material fósil de sitios arqueológicos como los que fueron reportados en las ciudades de Harappa o Mohenjo-daro, y con los marcadores presentados en este estudio, sería posible precisar el período en el que emergió *VRN-D4* y estudiar su relación con los hábitos culturales de la civilización del Valle Indio.

*VRN-D4* ha sido caracterizado como el principal alelo de hábito de crecimiento primaveral en *T. aestivum* ssp. *sphaerococcum*, y los otros alelos están presentes en muy baja frecuencia o completamente ausentes (Tabla 3.3). Por lo tanto es posible que el origen del alelo dominante de *VRN-D4* haya estado asociado con largas diferencias en la ocurrencia de la floración, de una forma similar a la que ha sido reportada en este estudio. Los análisis actuales de la frecuencia de distribución de los alelos dominantes de vernalización han revelado grandes diferencias entre diferentes ambientes lo que sugiere fuertes efectos de selección (Stelmakh 1998 y 1990; y Zhang et al. 2008).

Los primeros estudios de *T. aestivum* ssp. *sphaerococcum* sugirieron que esta subespecie difiere de ssp. *aestivum* solamente en un número pequeño de loci (58, 59). Sin embargo, nuestros resultados basados en un análisis completo del genoma sugiere que *T. aestivum* ssp. *sphaerococcum* es una de las subespecies más diferenciadas de *T. aestivum* (Figura 3.16 y Figura 3.17.D). Este resultado sugiere que *T. aestivum* ssp. *sphaerococcum* puede ser un recurso altamente valuable, en términos de nueva diversidad genética, para el mejoramiento de trigo hexaploide. Esto es consistente con

estudios previos que han sugerido a *T. aestivum* ssp. *sphaerococcum* como una nueva fuente de variabilidad en genes de altura de planta (*Rht*) (Mori, 2013; Percival, 1921) y de resistencia a roya amarilla (Maccaferri, 2015).

Nuestra hipótesis actual es que a medida que el cultivo de trigo se trasladó hacia el este y el oeste del área original de domesticación en la costa sur del Mar Caspio, se diferenció y adaptó a diferentes condiciones medioambientales. Las divergencias entre *T. aestivum* ssp. *sphaerococcum* y las otras subespecies occidentales de *T. aestivum* trazan un paralelo con la distribución de los lenguajes Indio-Europeos, donde las familias de lenguaje Indio-Iraníes son más diferentes que las familias occidentales como son la Celta, Germánica, Itálica y Baltico-Slavo (Bouckaert et al. 2012).

### **3.4.4 Nuevos mecanismos de control del requerimiento de vernalización**

De acuerdo a la información de secuencia obtenida en este estudio *VRN-D4* no presenta ninguno de los polimorfismos que han sido asociados anteriormente alelos primaverales en *VRN1* (Chu et al. 2011, Fu et al. 2005 y Yan et al. 2004b). Por lo tanto es necesario un mecanismo alternativo para explicar el hábito de crecimiento primaveral asociado con *VRN-D4*. Se proponen a continuación tres mecanismos no mutuamente excluyentes:

*Mecanismo 1 “Localización cromosómica”:* El gen *VRN-A1* está localizado ancestralmente en el brazo largo del cromosoma 5A en una región con una frecuencia de recombinación típica, mientras que *VRN-D4* está localizado en una región más proximal donde, hasta ahora, no se ha detectado recombinación (Esta tesis, capítulo 1). Está bien establecido que translocaciones cromosómicas y re-arreglos génicos pueden tener un gran impacto en la expresión génica. Además, genes vecinos pueden tener una alta correlación en perfiles de expresión génica (Williams et al. 2004 y Zhan et al. 2005) y pueden impactar la expresión de los genes en esta nueva localización. Sin embargo, no se han identificado genes cerca de *VRN-D4* por lo que el efecto potencial mediante este mecanismo, es a nuestro entender, limitado para explicar los efectos en la expresión de

*VRN-D4*. Los efectos de la posición cromosómica en la expresión génica también pueden estar asociados con la proximidad a regiones de eucromatina o heterocromatina. En *Drosophila*, estos efectos se pueden diseminarse hasta distancias considerables, pero comúnmente son los genes localizados cerca de los puntos de ruptura los más afectados (Lima-de-Faria 1983). Sin embargo, en el *locus* de *VRN-D4*, el gen está todavía rodeado de grandes fragmentos de la cromatina original del cromosoma 5AL (~80 kpb río arriba y ~200 kpb río abajo de *VRN-D4*), lo cual puede moderar los potenciales efectos de la diferente localización cromosómica en la expresión de *VRN-D4*.

*Mecanismo 2 “Efecto en cis del polimorfismo K123Q”*: El cambio de nucleótido A367C presente en *VRN-D4* resulta en un cambio de aminoácido (K123Q) que no ha sido detectado en ninguna de las secuencias de proteínas de *VRN1* disponibles hasta ahora en las bases de datos de secuencias actuales. Por lo tanto sería probable que este polimorfismo contribuya a la expresión temprana de *VRN-D4*. Sin embargo, debido a que *VRN-D4* es el único de los genes *VRN1* que está regulado positivamente durante los estadios tempranos del desarrollo en la línea TDF (Figura 3.1.E y Figura 3.2), cualquier efecto que involucre este cambio de aminoácido debe actuar en “*cis*”. Ya que cualquier mecanismo que involucre la regulación de algún otro gen intermediario (“*trans*”) debería afectar la expresión de los otros homólogos de *VRN1*. Aunque este argumento no invalida el potencial efecto del cambio de nucleótido A367C en la expresión temprana de *VRN-D4*, hace esta hipótesis poco probable.

*Mecanismo 3 “Mutaciones puntuales en la región RIP-3”*: Una explicación alternativa para la expresión temprana de *VRN-D4* es la presencia de tres cambios de nucleótido en la región RIP-3 del primer intrón de *VRN-D4*. Estos tres SNPs están localizados en un bucle de ARN que forma el sitio de unión de la proteína TaGRP2 (Figura 3.13C). La región RIP-3 en trigo es similar al sitio de unión de la proteína de Arabidopsis homóloga GRP7 (Figura 3.11.A), y uno de estos tres SNPs presentes en RIP-3<sup>*VRN-D4*</sup> está en una posición que ha sido descrita como crítica para la unión de GRP7 en Arabidopsis (Leder et al. 2014). La variación en GRP7 afecta la floración en Arabidopsis mediante la regulación de la expresión del gen del tipo MADS-box *FLOWERING LOCUS C* (*FLC*) (Streitner et al. 2008).

Estudios previos en trigo han mostrado que la sobreexpresión de *TaGRP2* retrasa la ocurrencia de la floración mientras que el silenciamiento mediante ARNi la acelera,



confirmando que GRP2 funciona como un represor floral (Xiao et al. 2014). Nuestros resultados de interacción ARN-proteína (REMSA) muestran que la presencia de los tres cambios de nucleótido presentes en la región RIP-3<sup>VRN-D4</sup> son suficientes para impedir la unión entre TaGRP2 y su sitio blanco (Figura 3.13). La unión de TaGRP2 al pre-ARNm de *VRN1* es requerida para la inhibición de la expresión de *VRN1*, por lo tanto la disrupción de la interacción puede explicar la expresión temprana de *VRN-D4*. Este mecanismo es también consistente con las deleciones del primer intrón de *VRN1* que incluyen la región RIP-3 y están asociados con hábito de crecimiento primaveral (Fu et al. 2005 y Hemming et al. 2009).

Estos tres SNPs presentes en la región RIP-3 de *VRN-D4* fueron también encontrados en la región RIP-3 de *VRN1* en trigo hexaploide de las variedades Jagger, Claire y Chinese Spring (Figura 3.11.B). El alelo de *VRN1* presente en Chinese Spring no ha sido caracterizado en detalle aún, pero las variedades Jagger y Claire han sido reportadas como invernales con un reducido requerimiento de vernalización (Díaz et al. 2013 y Li et al. 2013). Esta reducción en el requerimiento de vernalización fueron mapeadas en las dos variedades en el locus *VRN-A1*, pero el mecanismo causal es todavía controversial. La reducción del requerimiento de vernalización en Claire fue atribuída a una disminución en el número de copias de *VRN-A1* y en Jagger a un cambio de aminoácido en *VRN-A1* (A180V) (Li et al. 2013). La identificación de RIP-3<sup>VRN-D4</sup> en los alelos de Claire y Jagger provee una explicación alternativa para la reducción en el requerimiento de vernalización. Proponemos que la disminución de la unión del represor TaGRP2 al alelo RIP-3<sup>VRN-D4</sup> presente en estas dos variedades resulta en un aumento de los niveles de expresión del transcripto de *VRN-A1* reduciendo así el requerimiento de vernalización. Esta interpretación es consistente con los niveles de expresión reportados en estas dos variedades. Los niveles de expresión del transcripto de *VRN-A1* se incrementan más rápidamente en Jagger que en la línea invernal 2174 (RIP-3<sup>VRN-A1</sup>) (Li et al. 2013), lo mismo ocurre en Claire con respecto a las variedades invernales Malacca y Hereward (Díaz et al. 2012).

Otro aspecto que necesita ser señalados es que un corto tratamiento de vernalización de 3 semanas acelera la floración en Jagger y Claire por más de un mes (Li et al. 2013 y Díaz et al. 2013), pero no tiene un efecto significativo en TDF (*VRN-D4*) (Yoshida et al. 2010). Esto sugiere que otros factores adicionales podrían contribuir a esta diferencia de respuesta a la vernalización de RIP-3<sup>VRN-D4</sup> en TDF y en los alelos de *VRN-A1* en Jagger

y Claire. La diferente localización cromosómica o el cambio de nucleótido K123Q a nivel de proteína presente en *VRN-D4* pueden contribuir a estas diferencias, pero además también es posible que los diferentes fondos genéticos de estas variedades modulen el efecto del alelo  $\text{RIP-3}^{\text{VRN-D4}}$ . Por ejemplo, la variedad primaveral Chinese Spring puede ser interpretada como el efecto combinado de  $\text{RIP-3}^{\text{VRN-D4}}$  en *VRN-A1* y el alelo primaveral producto de la delección del primer intrón en *VRN-D1* también presente en esta variedad.

En los alelos de *VRN-A1* caracterizados se produjo una observación interesante, se encontró que la región RIP-3 de *T. urartu* así como las líneas cultivadas y salvajes secuenciadas todas poseen el SNP T2783C (alelo  $\text{RIP-3}^{\text{VRN-A1}}$ ). Los resultados aquí presentados de interacción RNA-proteína mostraron que esta mutación es suficiente para debilitar la interacción entre TaGRP2 y su secuencia de RNA blanco, aunque esta disrupción no fue tan fuerte como la observada para los tres SNPs presentes en el alelo  $\text{RIP-3}^{\text{VRN-D4}}$ . Es posible especular que el SNP T2783C en  $\text{RIP-3}^{\text{VRN-A1}}$  puede contribuir a la mayor expresión observada en *VRN-A1* con respecto a *VRN-B1* y *VRN-D* después de seis semanas de vernalización (Loukoianov et al. 2005).

Para corroborar esta hipótesis se deberían generar la combinación apropiada de alelos en un mismo fondo genético a modo de poder disectar el efecto propio de cada alelo. Por ejemplo se podría combinar el alelo presente en *VRN-A1* de la variedad Chinese Spring ( $\text{RIP-3}^{\text{VRN-D4}}$ ) con el alelo presente en TDC ( $\text{RIP-3}^{\text{VRN-A1}}$ ), estas dos líneas únicamente presentan polimorfismos a nivel de la región RIP-3. De la misma manera, la única fuente del alelo ancestral de la región RIP-3 parece ser el trigo diploide *T. monococcum*, que fue la secuencia que mostró una unión más fuerte con TaGRP2 en todas las condiciones estudiadas. Sería interesante transferir el alelo de *VRN-A1* de *T. monococcum* a un fondo genético hexaploide, por ejemplo Chinese Spring, y poder así comparar los efectos de los tres SNPs en la región RIP-3 en contraste con la secuencia canónica de este sitio.

### 3.5 CONCLUSIÓN

Este capítulo presenta distintas líneas de evidencia para corroborar que *VRN-D4* es el resultado de la inserción de un fragmento de ADN de aproximadamente 290 Kpb del

cromosoma 5AL incluyendo el gen de floración *VRN-A1* al brazo cromosómico 5DS (el único gen identificado en esta inserción). En primer lugar se mostró que esta copia adicional de *VRN-A1* se encuentra completamente ligada a *VRN-D4* en una población de mapeo que incluye 3.182 cromosomas recombinantes. Además, *VRN-D4* está ligado con una alta expresión de *VRN1* en estadíos tempranos de desarrollo en ausencia de vernalización. Esta copia adicional localizada en el cromosoma 5DS posee modificaciones específicas en una región regulatoria del primer intrón que se encontraron ligadas con el hábito de crecimiento en las distintas poblaciones utilizadas. La generación de mutaciones químicas inducidas en la copia de *VRN-A1* insertada en el cromosoma 5D resultó en plantas con floración tardía, demostrando que este gene es *VRN-D4*. El estudio de la frecuencia con la que *VRN-D4* se encuentra presente en diferentes subespecies de *T. aestivum* identificó a *T. aestivum* subsp. *sphaerococcum* como la fuente más probable de *VRN-D4*. Este resultado explica también la mayor frecuencia de *VRN-D4* en variedades de trigo de Asia. Finalmente, la caracterización de la región RIP3 en el primer intrón reveló nuevos mecanismos de la regulación del requerimiento de vernalización.

## CAPÍTULO

## 4

**Discusión general y conclusiones**

## 4.1 DISCUSION GENERAL

### 4.1.1 Discusión de los objetivos alcanzados

En el desarrollo del trabajo presentado en esta tesis se cumplieron con los objetivos planteados inicialmente como así también se expandieron a análisis más detallados en algunos aspectos no considerados inicialmente.

Los resultados del mapeo fino de la región candidata (Objetivo 1) mostraron que no fue posible detectar un candidato para *VRN-D4* mediante la estrategia tradicional de clonado posicional, lo que concuerda con los resultados finales de identificación de *VRN-D4* como un evento exclusivo de trigo hexaploide. Las regiones centroméricas, además de poseer un bajo contenido de polimorfismos, son regiones de baja recombinación genética (Lukaszewski 1993, Akhunov et al. 2003). En este trabajo se logró resolver la falta de polimorfismo mediante la utilización de una línea de sustitución del cromosoma 5D, lo que permitió incrementar la tasa de polimorfismo en el cromosoma 5D y facilitar el desarrollo de nuevos marcadores. Sin embargo, este incremento en el polimorfismo trajo aparejada una disminución en la frecuencia de recombinación (Dubcovsky et al. 1995). Esta reducción, junto con la localización de *VRN-D4* cerca de la región centromérica del cromosoma 5D, imposibilitó el clonado de *VRN-D4* mediante mapeo posicional tradicional. Para determinar la localización de *VRN-D4* a nivel de brazo cromosómico fue necesario separar los brazos cromosómicos 5DS and 5DL generando telocéntricos para el cromosoma 5D del padre portador del gen *VRN-D4*.

Dadas las limitaciones en la recombinación en la región del gen *VRN-D4*, se utilizaron estrategias distintas del clonado posicional tradicional. En primer lugar se caracterizó en detalle la expresión de los genes que regulan el requerimiento de vernalización en la línea portadora del alelo dominante de *VRN-D4* (Objetivo 3). Estos estudios mostraron la existencia de una copia adicional del gen de vernalización *VRN-A1* que fue mapeada en la región de *VRN-D4*. Estos resultados en combinación con la observación que esta nueva copia de *VRN-A1* en el cromosoma 5D es la única que se expresa temprano en el desarrollo en ausencia de vernalización sugirieron que la copia del gen de vernalización *VRN1* localizada en el brazo corto del cromosoma 5D era un buen gen candidato para *VRN-D4*. Para comprobar esta hipótesis se construyó un mapa físico de esta región

(Objetivo 2), mediante la secuenciación de BACs del brazo corto del cromosoma 5D de la línea ditelocéntrica del brazo cromosómico 5DS en la variedad Chinese Spring. El análisis de esta región demostró la inserción de un segmento de ~290 kpb del cromosoma 5AL incluyendo el gene *VRN-A1*. Mediante la utilización de mutaciones inducidas se demostró que esta copia adicional del gen de vernalización *VRN-A1* es *VRN-D4* (Objetivo 5).

Los objetivos planteados en esta tesis que incluían la detección y validación del gen candidato, así como el estudio de la expresión en relación a los genes *VRN1/VRN2/VRN3* fueron cumplidos. Mediante la combinación de distintas estrategias se mapeo con precisión el gen *VRN-D4*, se describió la región física flanqueando al gen, y se validó el gen candidato utilizando mutantes. Finalmente este estudio se expandió a la caracterización de la distribución geográfica de este gen y de los posibles mecanismos de regulación.

#### 4.1.2 Análisis de las hipótesis planteadas

**Hipótesis 1:** “*VRN-D4* es un gen diferente a los genes de vernalización clonados anteriormente, y mutaciones en este gene afectaran las respuesta a vernalización en trigo y la duración del “*periodo a espigazón*”.”

Esta hipótesis es aceptada. A pesar de que se propuso a *VRN-D4* como un nuevo gene involucrado en la regulación génica de la respuesta de vernalización, los resultados aquí mostrados indican que *VRN-D4* es el producto de la duplicación de un gen descrito previamente, pero insertado en una región cromosómica diferente y con variaciones en su secuencia que afectan su expresión y afectan la floración. La caracterización de *VRN-D4* sugiere nuevos mecanismos para la regulación del requerimiento de vernalización. El alelo descrito en *VRN-D4* muestra la interacción más débil entre la región reguladora del primer intrón y la proteína represora TaGRP2. En base a estos resultados se propone la hipótesis que este cambio en la interacción con TaGRP2 es responsable, al menos en parte, del reducido requerimiento de vernalización. La variación en la región regulatoria RIP-3 merece ser estudiada en detalle ya que podría aportar una nueva herramienta para manipular el requerimiento de vernalización en las variedades invernales.

El alelo de RIP-3<sup>VRN-A1</sup> (SNP T2783C) se detectó en todas las variedades estudiadas tanto de trigos tetraploides como hexaploides y se demostró que interactúa más débilmente con el represor GRP2 que el alelo que contienen la secuencia canónica (únicamente detectada en *T. monococcum*). Por lo tanto, sería interesante estudiar el efecto del alelo de *T. monococcum* en las variedades invernales, o alternativamente editar este sitio en variedades modernas, por ejemplo mediante la utilización del sistema CRISP/Cas9 (Wang et al. 2014). En esta tesis se propone la hipótesis que el alelo de *T. monococcum* podría estar asociado a un aumento en el requerimiento de vernalización. Este alelo es nuevo en trigo pan y pasta, y presenta una oportunidad de desarrollar nuevas adaptaciones a zonas con inviernos largos.

Con los marcadores desarrollados en esta tesis es posible desarrollar líneas isogénicas para testear los efectos de los diferentes alelos de RIP3. Por ejemplo, es posible combinar mediante cruzamientos las variantes de los alelos RIP-3<sup>VRN-A1</sup> (presente en la línea TDC) y RIP-3<sup>VRN-D4</sup> (presente en la variedad Chinese Spring) para estudiar sus efectos en la duración del requerimiento de vernalización necesario para saturar esta respuesta. En este caso se esperaría encontrar que los tres SNPs presentes en RIP-3<sup>VRN-D4</sup> estén ligados con una disminución en el requerimiento de vernalización, y una floración más temprana en tratamientos de vernalización limitados (por ejemplo, 3 semanas).

**Hipótesis 2:** “En el modelo de regulación de la etapa reproductiva de trigo, mediada por respuesta a vernalización, *VRN-D4* es reprimido, directa o indirectamente por *VRN2*, y su expresión promueve la expresión de *VRN1*.”

Esta hipótesis es rechazada debido a que *VRN-D4* no este regulado por vernalización. Como se ha mostrado en el último capítulo, *VRN-D4* se expresa temprano en el desarrollo aun en las ausencia de vernalización, y diferentes tratamientos de vernalización no aceleran la floración cuando *VRN-D4* está presente (Yoshida et al. 2010).

**Hipótesis 3:** “La diversidad alélica de *VRN-D4* se encontrará asociada con variaciones en la secuencia de sus regiones regulatorias (promotor y/o intrón).”

La hipótesis se propuso en base al carácter dominante de *VRN-D4* y es aceptada. La hipótesis más probable es que el carácter primaveral de *VRN-D4* este determinado por las diferencias en la región regulatoria RIP-3 del primer intrón. Sin embargo esto no es fácil de demostrar dado que *VRN-D4* es un evento único, y no es posible estudiar

diversidad alélica en este locus de forma tradicional. Para demostrar esta hipótesis hemos iniciado la edición del sitio RIP-3 utilizando CRISPR-Cas9.

**Hipotesis 4:** “La relaciones de homeología existentes entre los genomas de trigo (A, B y D), permitirá identificar genes ortólogos a *VRN-D4* en los genomas A y B, y caracterizar la variabilidad existente en los mismos para ampliar las posibilidades de manipulación del desarrollo de la etapa reproductiva en programas de mejoramiento.”

Esta hipótesis es rechazada. *VRN-D4* se originó mediante una inserción de un fragmento de ADN del cromosoma 5A (290kb) al cromosoma 5D, después del origen del trigo hexaploide. Por lo tanto, este evento no se encuentra presente en ninguno de los genomas ancestrales de trigo, y por lo tanto no hay genes homeólogos a *VRN-D4* en los genomas A y B. Esto se demostró por análisis BLAST de las secuencias genómicas de los brazos cromosómicos 5AS y 5BS. Ninguna secuencia homóloga a *VRN-D4* fue detectada en estos brazos cromosómicos.

. Esto también implica que este evento ocurrió en tiempos recientes y muy probablemente se encuentre relacionado con la selección de trigo de floración más temprana en la zona del Punjab en India

## 4.2 CONCLUSIÓN GENERAL

A continuación se listan los aportes particulares y los recursos generados en esta tesis:

- Se desarrolló un mapa fino de la región centromérica del cromosoma 5D, que constituye el mapa más detallado de esta región. Este mapa es de utilidad para los proyectos actuales de secuenciación de trigo pan ya que puede ser utilizado para la construcción de mapas físicos detallados de la región.
- Se comprobó la localización del centrómero activo del cromosoma 5D en una región diferente al centrómero activo de los cromosomas colineares en *B. distachyon* y (Cr. 4) y arroz (Cr 12).
- Se descartó a *TaVIL-D1* como gen candidato para *VRN-D4*. En base a este resultado se reiniciaron estudios de mutantes de *TaVIL-D1* para entender su función.
- Se determinó el orden relativo entre los genes de trigo presentes en la región centromérica del cromosoma 5D y los genes de otras especies modelo (arroz y *Brachypodium*), lo que facilitará futuros estudios comparativos de dicha región.



- Se mostró que el efecto de *VRN-D4* en tiempo a floración es menor que los alelos dominantes primaverales de *VRN-A1*, *VRN-D1* y *VRN-B3*, y similar a *VRN-B1*.
- Se demostró que el efecto de floración temprana de *VRN-D4* es epistático a los alelos invernales de *VRN1/VRN2/VRN3*.
- Se determinó que *VRN-D4* es el resultado de la inserción de una copia de *VRN-A1* en el cromosoma 5D.
- Se desarrollaron marcadores moleculares para *VRN-D4* que pueden ser fácilmente utilizados para:
  - La selección de este alelo e incorporación en programas de mejoramiento.
  - Estudiar el efecto en tiempo a floración de *VRN-D4* en diferentes fondos genéticos.
  - Analizar la frecuencia de este alelo en diferentes poblaciones.
  - El estudio de ADNs de muestras arqueológicas lo que permitirá entender el origen y la selección de *VRN-D4* en la India.
- Se presentó evidencia que sugiere que *VRN-D4* se originó en *T. aestivum* spp. *sphaerococcum*, y se plantearon hipótesis sobre la importancia de este gen en el cultivo de trigo en la región del Punjab.
- Se caracterizó la diversidad alélica de *T. aestivum* spp. *sphaerococcum* y se comparó con la diversidad en las otras especies de *T. aestivum*.
- Se demostró que *T. aestivum* spp. *sphaerococcum* es una subespecie grupo muy diferenciada genéticamente de las otras subespecies y por lo tanto es sugerida como una nueva fuente de diversidad genética para el mejoramiento genético de variedades modernas.
- Se generó una población de 1153 líneas de TDF mutagenizadas con EMS. Esta población es de libre disponibilidad para la comunidad científica y puede ser utilizada para estudios de genómica funcional en trigo hexaploide.
- Se demostró el efecto de la variación alélica en la región RIP3 en la capacidad de esta región de interactuar con la proteína GRP2.
- Se identificaron variedades de trigo hexaploides que tienen una secuencia RIP3 similar a la de *VRN-D4* en el gene *VRN-A1*, y se mostró que este alelo de *VRN-A1* está asociada a un requerimiento reducido de vernalización.

- Los nuevos alelos identificados en *VRN-A1* y *VRN-D4* representan una nueva herramienta para manipular el ciclo de crecimiento de trigo en los programas de mejoramiento.

## BIBLIOGRAFÍA

- Adzhubei I, Schmidt S, Peshkin L, Ramensky V, Gerasimova A, Bork P, Kondrashov A and Sunyaev S (2010). A method and server for predicting damaging missense mutations. *Nat Methods* 7:248-249.
- Akhunov ED et al. (2003). The Organization and Rate of Evolution of Wheat Genomes Are Correlated With Recombination Rates Along Chromosome Arms. *Genome Research*, 13(5), 753–763.
- Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo H, Huo N, Lazo GR, Luo MC, Ma YQ, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, Dvorak J (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* 11:702.
- Alghabari F, Lukac M, Jones H, Gooding M (2014). Effect of Rht alleles on the tolerance of wheat grain set to high temperature and drought stress during booting and anthesis. *J. Agron. Crop Sci.* 200:36–45.
- Allouis S, Moore G, Bellec A, Sharp R, Faivre Rampant P, Mortimer K, Pateyron S, Foote TN, Griffiths S, Caboche M, Chalhoub B (2003). Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm 'Chinese Spring'. *Cereal Res Commun* 31:331-338.
- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993). Optimizing parental selection for genetic-linkage maps. *Genome* 36:181-186.
- Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, Gray RD, Suchard MA and Atkinson QD (2012). Mapping the Origins and Expansion of the Indo-European Language Family. *Science* 337:957-960.
- Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005). Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* 17:343-360.

- Chen A and Dubcovsky J (2012). Wheat TILLING mutants show that the vernalization gene *VRN1* down-regulates the flowering repressor *VRN2* in leaves but is not essential for flowering. *PLoS Genet* 8:e1003134.
- Chen A, Li C, Hu W, Lau M, Lin H, Rockwell N, Martin S, Jernstedt J, Lagarias C, Dubcovsky J (2014). PHYTOCHROME C plays a major role in the acceleration of wheat flowering under long-day photoperiod. *Proc Natl Acad Sci USA* 111:10037-10044.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7.
- Chu C, Tan C, Yu G, Zhong S, Xu S, and Yan L (2011). A Novel Retrotransposon Inserted in the Dominant *Vrn-B1* Allele Confers Spring Growth Habit in Tetraploid Wheat (*Triticum turgidum* L.). *G3* 7: 637–645.
- Danyluk J, et al. (2003). *TaVRT-1*, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiol* 132:1849-1860.
- Devos K, Sorrells M, Anderson J, Miller T, Reader S, Lukaszewski A, Dubcovsky J, Sharp P, Faris J and Gale M(1999). Chromosome aberrations in wheat nullisomic-tetrasomic and ditelosomic lines. *Cereal Res Commun* 27:231-239.
- Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature* 418:700-707
- Diaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012). Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS One* 7:e33234.
- Distelfeld A, Li C, Dubcovsky J (2009). Regulation of flowering in temperate cereals. *Curr Opin Plant Biol* 12:178-184.
- Distelfeld A, Tranquilli G, Li C, Yan L, Dubcovsky J (2009b). Genetic and molecular characterization of the *VRN2* loci in tetraploid wheat. *Plant Physiol* 149:245-257.
- Draper J, Mur LA, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge AP (2001). *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiol* 127:1539-1555.

- Dubcovsky J, Luo M, Dvorak J (1995). Differentiation between homoeologous chromosomes 1A of wheat and 1Am of *Triticum monococcum* and its recognition by the wheat *Ph1* locus. *Proc Natl Acad Sci USA* 92:6645-6649.
- Dubcovsky J and Dvorak J (2007). Genome Plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316:1862-1866.
- Earl DA, Vonholdt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359-361.
- Endo TR, Gill BS (1996). The deletion stocks of common wheat. *J Hered* 87:295-307.
- Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform* 1:47-50.
- Fu D, Szűcs P, Yan L, Helguera M, Skinner J, von Zitzewitz J, Hayes P, Dubcovsky J (2005). Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Mol Genet Genomics* 273:54-65.
- Godfray C , Beddington J, Crute I, Haddad L, Lawrence D, Muir J, Pretty J, Robinson S, Thomas S, and Toulmin C (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science* 327:812-818.
- Goncharov NP, Shitova IP (1999). The inheritance of growth habit in old local varieties and landraces of hexaploid wheats. *Russian J Genet* 35:386-392.
- Goncharov NP (2003). Genetics of growth habit (spring vs winter) in common wheat: confirmation of the existence of dominant gene *Vrn4*. *Theor Appl Genet* 107:768-772.
- Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006). Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature* 439:749-752.
- Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL (2008). The Vienna RNA websuite. *Nucleic Acids Res* 36:W70-W74.
- Hemming MN, Peacock WJ, Dennis ES, Trevaskis B (2008). Low-temperature and daylength cues are integrated to regulate *FLOWERING LOCUS T* in barley. *Plant Physiol* 147:355-366.

- Hutchinson J, Allchin FR, Vishnu-Mittre (1976). India; local and introduced crops. *Phil Trans R Soc Lond, Series B* 275:129–141.
- Iwaki K, Haruna S, Niwa T, Kato K (2001). Adaptation and ecological differentiation in wheat with special reference to geographical variation of growth habit and *Vrn* genotype. *Plant Breeding* 120:107-114.
- Iwaki K, Nakagawa K, Kuno H, Kato K (2000). Ecogeographical differentiation in east Asian wheat, revealed from the geographical variation of growth habit and *Vrn* genotype. *Euphytica* 111:137-143.
- Jia J et al. (2013). *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496:91-95.
- Jiang N, Bao ZR, Zhang XY, Eddy SR, Wessler SR (2004). Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431:569-573.
- Kato K, Yamashita M, Ishimoto K, Yoshino H, Fujita M (2003). Genetic analysis of two genes for vernalization response, the former *Vrn2* and *Vrn4*, using PCR based molecular markers. In: Pogna NE, Romano N, Pogna EA, Galterio G (eds) *Proceeding of 10<sup>th</sup> international wheat genetics symposium*. Inst Sperimentale per la Cerealcolture, Paestum, Italy, pp 971–973
- Knott DR (1959). The inheritance of rust resistance. IV. Monosomic analysis of rust resistance and some other characters in six varieties of wheat including Gabo and Kenya Farmer. *Can J Plant Sci* 39:215–228.
- Krasileva KV, Buffalo V, Bailey P, Pearce S, Ayling S, Tabbita F, Soria M, Wang S, Consortium I, Akhunov E, Uauy C, Dubcovsky J (2013). Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology*, 14:R66.
- Laurie, D, Pratchett N, Bezant J and Snape J (1995). RFLP mapping of 5 major genes and 8 quantitative trait loci controlling flowering time in a winterxspring barley (*Hordeum-vulgare* L) cross. *Genome* 38:575–585.
- Law CN, Stuka J, Worland AJ. (1978). A genetic study of day-length response in wheat. *Heredity* 41:185–191.
- Leder V, Lummer M, Tegeler K, Humpert F, Lewinski M, Schüttpelz M and Staiger D (2014). Mutational definition of binding requirements of an hnRNP-like protein

in Arabidopsis using fluorescence correlation spectroscopy. *Biochem Bioph Res Co* 453:69-74.

Lima-de-Faria A (1983). *Molecular evolution and organization of the chromosome* (Elsevier, Amsterdam).

Ling HQ, et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496:87-90.

Linkiewicz AM, Qi LL, Gill BS, Ratnasiri A, Echalié B, Chao S, Lazo GR, Hummel DD, Anderson OD, Akhunov ED, Dvorak J, Pathan MS, Nguyen HT, Peng JH, Lapitan NL, Miftahudin, Gustafson JP, La Rota CM, Sorrells ME, Hossain KG, Kalavacharla V, Kianian SF, Sandhu D, Bondareva SN, Gill KS, Conley EJ, Anderson JA, Fenton RD, Close TJ, McGuire PE, Qualset CO, Dubcovsky J (2004). A 2500-locus bin map of wheat homoeologous group 5 provides insights on gene distribution and colinearity with rice. *Genetics* 168:665-676.

Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J (2005). Regulation of VRN-1 vernalization genes in normal and transgenic polyploid wheat. *Plant Physiol* 138:2364-2373.

Lukaszewski AJ, Gustafson JP (1983). Translocations and modifications of chromosomes in triticales X wheat hybrids. *Theor Appl Genet* 64:239-248.

Lukaszewski AJ, Curtis CA (1993). Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theor Appl Genet* 86:121-127.

Luo MC, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, Jorgensen CM, Zhang Y, McGuire PE, Pasternak S, Stein JC, Ware D, Kramer M, McCombie WR, Kianian SF, Martis MM, Mayer KF, Sehgal SK, Li W, Gill BS, Bevan MW, Simkova H, Dolezel J, Weining S, Lazo GR, Anderson OD, Dvorak J (2013). A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proc Natl Acad Sci USA* 110: 7940–7945.

Maccaferri, M., J. Zhang, P. Bulli, Z. Abate, S. Chao, D. Cantu, E. Bossolini, X. Chen, M. Pumphrey, and J. Dubcovsky. (2015). A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). *G3* 5:449-465.

- Marcellos H, Single W (1984). Frost injury in wheat ears after ear emergence. *Aust J Plant Physiol* 11:7–15.
- Mayer KFX, et al. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345.
- Maystrenko OI (1980). Cytogenetic study of the growth habit and ear-emergence time in wheat (*Triticum aestivum* L.). In: *Well-Being of mankind and genetics*, vol. 1, book 2. Proc 14th Int Congress of Genetics. MIR, Moscow, Russia, pp 267–282.
- Mori N, Ohta S, Chiba H, Takagi T, Niimi Y, Shinde V, Kajale M, Osada T (2013). Rediscovery of Indian dwarf wheat (*Triticum aestivum* L. ssp. *sphaerococcum* (Perc.) MK.) an ancient crop of the Indian subcontinent. *Genet Resour Crop Ev* 6:1771-1775.
- Murphy RL, Morishige DT, Brady JA, Rooney WL, Yang S, Klein PE, and Mullet JE (2014). *Ghd7* (Ma6) represses sorghum flowering in long days: *Ghd7* alleles enhance biomass accumulation and grain production. *Plant Genome US* 7:2.
- Nevo E, Beiles A (1989). Genetic diversity of wild emmer wheat in Israel and Turkey - structure, evolution, and application in breeding. *Theor Appl Genet* 77:421-455.
- Ng PC, Henikoff S (2001). Predicting deleterious amino acid substitutions. *Genome Res* 11:863-874.
- Nitcher R, Distelfeld A, Tan C, Yan L, Dubcovsky J (2013). Increased copy number at the *HvFT1* locus is associated with accelerated flowering time in barley. *Mol Genet Genomics*. 288:261-275.
- O'Brien L MM, Wrigley C, Appels R (2001). Genetic pool of Australian wheats. In: Bonjean AP, Angus WJ (eds) *The world wheat book*. Lavoisier, Paris, pp pp 611-648.
- Parry, M. A. J., M. Reynolds, M. E. Salvucci, C. Raines, P. J. Andralojc, and X. G. Zhu, et al. (2011). Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *J Exp Bot* 62:453–467.
- Paterson AH, Bowers JE, Chapman BA (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 101:9903-9908.



- Percival J (1921). *The Wheat Plant, a Monograph*. Duckworth and Co, London Ellerton 1939,
- Peleg Z, Fahima T, Abbo S, Krugman T, Nevo E, Yakir D and Saranga Y (2005). Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell Environ* 28:176-191.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Pugsley AT (1972). Additional genes inhibiting winter habit in wheat. *Euphytica* 21:547-552.
- Pugsley AT (1971). A genetic analysis of the spring-winter habit of growth in wheat. *Aust J Agric Res* 22:21-31.
- Pugsley AT (1972). Additional genes inhibiting winter habit in wheat. *Euphytica* 21:547-552.
- Purugganan, MD, and Fuller DQ (2009). The nature of selection during plant domestication. *Nature* 457: 843–8.
- Qi L, Friebe B, Wu J, Gu Y, Qian C, Gill BS (2010). The compact *Brachypodium* genome conserves centromeric regions of a common ancestor with wheat and rice. *Funct Integr Genomic* 10:477-492.
- Reynolds, M., D. Bonnett, S. C. Chapman, R. T. Furbank, Y. Manes, and D. E. Mather, et al. (2011). Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. *J Exp Bot* 62:439–452.
- Saini HS, Sedgley M, Aspinall D. (1983). Effect of heat stress during floral development on pollen tube growth and ovary anatomy in wheat (*Triticum aestivum* L.). *Aust J Plant Physiol* 10:137–144.
- Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Marin W (2002). Genetics and geography of wild cereal domestication in the Near East. *Nat Rev Genet* 3 420–441
- Scarth, R, and Law, C N. (1983). The location of the photoperiod gene, Ppd2 and an additional genetic factor for earemergence time on chromosome 2B of wheat. *Heredity* 51:607–619.

- Sears EM S-SL (1978). The telocentric chromosomes of common wheat. In: Ramanujam S (ed) Proc of 5th Int Wheat Genetics Symp, New Delhi, India pp 389–407.
- Semenov MA, Mitchell RAC, Whitmore AP, Hawkesford MJ, Parry MAJ, Shewry PR. (2012). Shortcomings in wheat yield predictions. *Nat Clim Change* 2:380–382.
- Stelmakh AF (1987). Growth habit in common wheat (*Triticum aestivum* L EM. Thell. *Euphytica* 36:513–519
- Stelmakh AF, Avsenin VI (1996). Alien introgression of spring habit dominant genes into bread wheat genomes. *Euphytica* 89:65–68.
- Streitner C, et al. (2008). The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in *Arabidopsis thaliana*. *Plant J* 56:239-250.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725-2729.
- Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES (2003). MADS box genes control vernalization-induced flowering in cereals. *Proc Natl Acad Sci USA* 100:13099-13104.
- Trevaskis B, Hemming MN, Dennis ES, Peacock WJ (2007). The molecular basis of vernalization-induced flowering in cereals. *Trends Plant Sci.* 12:352-7.
- Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006). A NAC gene regulating senescence improves grain protein, zinc and iron content in wheat. *Science* 314:1298-1300.
- Uauy C, Paraiso F, Colasuonno P, Tran R, Tsai H, Berardi S, Comai L, Dubcovsky J (2009). A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat *BMC Plant Biol* 9:115-128.
- Vara Prasad P, Djanaguiraman M. (2014). Response of floret fertility and individual grain weight of wheat to high temperature stress: sensitive stages and thresholds for temperature and duration. *Funct Plant Biol* 41:1261–1269.
- von Zitzewitz J, Szucs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, Casas A, Chen TH, Hayes PM, Skinner JS (2005). Molecular and structural characterization of barley vernalization genes. *Plant Mol Biol* 59:449-467.

- Wang S, et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 SNP array. *Plant Biotechnol J* 12:787-796.
- Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C and Qiu J (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* 32:947–951
- Weber SA (1999). Seeds of urbanism: palaeoethnobotany and the Indus Civilization. *Antiquity*, 73, 813-826.
- Weber S, Kashyap A, Harriman D (2010). Does size matter: the role and significance of cereal grains in the Indus civilization. *Archaeol Anthropol Sci* 2:35–43
- Weir BS (1996). *Genetic data analysis II* (Sinauer Publishers Sunderland, MA).
- Welsh JR et al. (1973). Genetic control of photoperiod response in wheat. *Proceedings of the 4th International Wheat Genetics Symposium* 879-884.
- Wicker T, Guyot R, Yahiaoui N, Keller B (2003). CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. *Plant Physiol* 132:52-63.
- Wicker T, Buchmann JP, Keller B (2010). Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res* 20:1229-1237.
- Worland A. J. (1996). The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica* 89, 49–57
- Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat Genet* 40:761-767.
- Xiao J, Xu S, Li C, Xu Y, Xing L, Niu Y, Niu Y, Huan Q, Tang Y, Zhao C, Wagner D, Gao C, Chong K (2014). O-GlcNAc-mediated interaction between VER2 and TaGRP2 elicits TaVRN1 mRNA accumulation during vernalization in winter wheat. *Nat Commun* 5:4572.
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003). Positional cloning of the wheat vernalization gene *VRN1*. *Proc Natl Acad Sci USA* 100:6263-6268.

- Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004a). The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303:1640-1644.
- Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J (2004b). Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theor Appl Genet* 109:1677-1686.
- Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006). The wheat and barley vernalization gene *VRN3* is an orthologue of FT. *Proc Natl Acad Sci USA* 103:19581-19586.
- Yoshida T, Nishida H, Zhu J, Nitcher R, Distelfeld A, Akashi Y, Kato K, Dubcovsky J (2010). *Vrn-D4* is a vernalization gene located on the centromeric region of chromosome 5D in hexaploid wheat. *Theor Appl Genet* 120:543-552.
- Yousafzai FK, Al-Kaff N, Moore G (2010). The molecular features of chromosome pairing at meiosis: the polyploid challenge using wheat as a reference. *Funct Integr Genomic* 10:147-156.
- Zhang XK, Xia XC, Xiao YG, Dubcovsky J, He ZH (2008). Allelic variation at the vernalization genes *Vrn-A1*, *Vrn-B1*, *Vrn-D1* and *Vrn-B3* in Chinese common wheat cultivars and their association with growth habit. *Crop Sci* 48:458-470.
- Zhu J, Pearce S, Burke A, See DR, Skinner DZ, Dubcovsky J and Garland-Campbell K. (2014). Copy number and haplotype variation at the VRN-A1 and central FR-A2 loci are associated with frost tolerance in hexaploid wheat. *Theor Appl Genet* 127:1183–1197.

## APÉNDICE

Tabla A.1. Marcadores utilizados para construir el mapa genético de la región de *VRN-D4*.

EST Trigo	Secuencia de cebadores (5' - 3')	Producto (pb)	Temperatura (°C)	Detección
<b>BE591275</b>	F- CCTGAGAGTGAACCAGAGACT R- ATACGCCTGGACTCCTCAGA	1098	62	<i>SacI</i>
<b>BQ167501</b>	F- CCATCCTCTTAATTGCGATAC R- CATCGAACAGATAAACGTAAG	808	60	Indel
<b>BE499257</b>	F- GGTTTGCGTTTGTGGAGTTT R- GCTACAGCATAAAAAGTTTCACAG	992	57	<i>HhaI</i>
<b>BE445181</b>	F- TATAGTGATTAGCAAATGATGGC R- CTTCTTACATCCACGGCAT	683	57	<i>HaeIII</i>
<b>BE404490</b>	F- GTTTCACACAAAGACGAGTAC R- CAACCCTTCTACTAACAGTA	876	56	Sec.
<b>BE606654</b>	F- ACCGACAAGAACTCCTAGAT R- CAAAAGCATCGCAGAGAAACAC	534	58	<i>BsaAI</i>
<b>BG263382</b>	F- ACCCATGTGTACAAAATCTTCC R- TCCAATCTCGTCCACAATGA	1921	64*	<i>BspHI</i>
<b>TaAGL31</b>	F- AAGCTGCGCCACCAGATCCA R- GGAACCGTCGAAGTTAATTTAATAGCCT	704	64*	<i>NsiI</i>
<b>BE403761</b>	F- CCGTCTGAATTATTTACAGTC R- CAGTGCCCGTGCCAATCTC	517	56	Sec.
<b>CJ717651</b>	F- CGGATAGTTCATACTTTTACTTG R- CCATCATCGATAATGCTCAT	1736	64*	Sec.
<b>CJ930261</b>	F- AGCGTTAGCCATCATCAAGC R- AGGCAAGGCAGAAAGGAAAT	1839	66*	Sec.
<b>CJ680715</b>	F- ATCTAGAATTACTATGGTAGGCAG R- GTAAGTTCTTTACCAAGGTCAA	1540	61*	Sec.
<b>CJ522104</b>	F- CTCTTGTAACCTTGCCAGCC R- TCACCAGATTACACCGTCCA	1960	67*	Sec.
<b>BJ315664</b>	F- ACTTGCGTTCCATGTTACACTTGTAGG R- GCGCTCCTGGCAGCAGTCTC	687	62*	<i>TaqI</i>
<b>CJ521028</b>	F- TG GTTGTGTACCTGCCTGAAAGTTAAGC R- CAGAAGGGAATTGTATTGTTTAAATGCAAATGGC	786	59*	<i>TaqI</i>
<b>TaFT-like</b>	F- CTCCATTGGTACGTGCTAAG R- GCGACTACACCCGCTTCTAC	1207	66*	<i>BglI</i>
<b>BE44702</b>	F- ACGATGTAAATGACTAGAAATCTGTCTGAT R- GCTTCATCACCAAGGTCATGTGG	210	60	<i>MnI</i>
<b>BE444353</b>	F- GCCGTTTGATGTTGGCACGA R- AGGGAATGTGACTGGAGGTG	922	63	<i>SspI</i>
<b>BE405060</b>	F- TGCACTTTCAGAATCGAAGCG R- GGGCTTCGCGTCTTCGTGA	454	60	<i>HhaI</i>

\*94°C for 5 m, 94°C por 30 s, 8 ciclos of iniciales (-0.5°C por ciclo), seguido de 40 ciclos de PCR, extensión final de 7 min a 72°C. El marcador BG313707 fue publicado previamente (Yoshida et al. 2010) y no fue incluido en esta tabla.

**Tabla A.2. Condiciones de PCR utilizadas para secuenciar el gen *VIL1* en trigo diploide y hexaploide.**

Región	Primer	Secuencia (5' - 3')	Producto (pb) <sup>a</sup>	Temperatura annealing (°C)	Annelaing (seg)	Extensión (seg)	Ciclos
<b>Trigo diploide</b>							
5'UTR to intron1	TmVIL1/5UTRF1	GAGGCCAATTATGGCTGCTA	1538	60	60	90	35
	TmVIL1/Ex2R	AACGTGGCAGGACGACTC					
Exon1to exon2	TmVIL1/WF2	ATGGAGTCGACCGGAGGA	1130	60	30	90	35
	TmVIL1/Ex2R	AACGTGGCAGGACGACTC					
Exon2 to exon3	TmVIL1/F2	CTCATCGGAGACTGGTGATA	719	57	60	90	38
	TmVIL1/Ex3R	AGGGGACCAACCTCAGTTTC					
Exon3 to exon4	TmVIL1/Ex3F1	ATAGGCTTTTGGATGGCACA	949	62	30	60	35
	TmVIL1/R3	TCTCATTTCGGAGGCACTCT					
Exon4 to 3'UTR	TmVIL1/F1	AGGTTGCTCGTCTACTGCCA	781	62	30	60	35
	TmVIL1/WR2	AGTGTTCCTATCCGGATT					
3'UTR	TmVIL1/IPF1	GTCTTGACCTTTATCCGCAC	1367	60	60	90	35
	TmVIL1/3UTRR1	GGCAGGTGGATTATTTGGTG					
<b>Trigo hexaploide</b>							
5'UTR to intron1	VIL1-up1060-F1	CCCCCTCTAGACGCCTACCTTG	1315 (1314)	65	60	90	35
	VIL-D1-int1-R3	CACGCAGCATGGGACCCACCCA					
Exon1 to intron1	VIL-1-ex1-F1	AGCGGAACCATCCGCTCTC	264 (264)	62.5	30	60	36
	VIL-D1-int1-R2b	GCGCGGGAGAAGAATGACCT					
Intron1to exon2	VIL-D1-int1-F1	CAGCTTGCTTATTAGGCTACCATT	656 (656)	57	60	60	33
	VIL-D1-ex2-R2	ACTGGCATCGTTTGATTCTAGC					
Intron1 to intron2	VIL-D1-T-int1-F1	GCATATTCAAGTGTTCAACCAGAGT	1001 (989)	55	30	60	30*
	VIL-D1-T-int2-R1	CAACAAAACCTATCACAGAAAAGTACC					
Exon2 to intron2	TmVIL1/Ex2F1	CTTGGGCAATCTATGCACCT	537 (527)	57	60	60	34
	VIL-D1-int2-R2	AAAGGTACCATAAATAAGCAAAGCA					
Intron2 to exon4	VIL-D1-T-int3-F1	TCTGTAAGAGCGTTTAGATCACCA	823 (823)	60	30	60	35
	TmVIL1/R2	GATCCGAAAGGCATACTCCG					
Entron3 to 3'UTR	VIL-D1-int3-F1	GTGGGTTCATGTGGATTCTTTT	1317 (1317)	60	30	60	38
	VIL-D1-T-3UTR-R1	CGAAGTGGA AAAACTGTGATACCT					
Exon4 to 3'UTR	TmVIL1/IPF1	GTCTTGACCTTTATCCGCAC	768 (767)	60	30	60	35
	VIL-D1-3UTR-R1	CCAGAATTTAGCGGTGAGAAGA					
3'UTR	VIL-D1-3UTR-F1	CACACAACAAAATGAAATGAGTGA	826 (828)	60	30	60	35
	TmVIL1/3UTRR1	GGCAGGTGGATTATTTGGTG					

a: Los tamaños de los productos de PCR fueron estimados utilizando las siguientes secuencias: Para trigo diploide, la secuencia de *Triticum monococcum* (DQ886922; Fu et al., 2007). Para trigo hexaploide, TDF (AB846583) y Hayakomugi (AB846584), y CS<sub>5402</sub> (AB846585) en paréntesis. \*94<sup>0</sup>C por 5 m, 94<sup>0</sup>C por 30 s, 10 ciclos iniciales (-0.5<sup>0</sup>C por ciclo), seguido de 30 ciclos de PCR, con una extensión final de 7 min a 72<sup>0</sup>C.



**Tabla A.3. Cebadores de PCR y sondas de ARN utilizadas.**

Descripción	Nombre	Secuencia
Amplificación de ADNc en <i>VRN-D4</i>	cDNA-VRN1-A-5DS-F cDNA-VRN1-A-5DS-R	GTTCTCCACCGAGTCATGTA TGTGGCTCACCATCCACG
Búsqueda en biblioteca de BACs 5DS	5P-BAC-F1 5P-BAC-R1 3P-BAC-F1 3P-BAC-R1 3P-BAC-F2 3P-BAC-R2	GCTCAAGAGGTAGCAATTCC CATTTTCTTATTGGGAACCGT CAAAATATGGAAGGGTGAAG TAGGTACCTGCACACACACA ACAACGTAGATGCACACACA CTCCATTGAATAGCTTGTCG
<i>VRN-D4</i> TILLING	TILL-5DS-F1 TILL-A-DS-R10	GATCTTGAATCTTTGAATCgCC* GCTGCAGCTTGCTACTTTAC
Detección del SNP A367C**	VA1-3PRIME-F2 VRN-A1-CAPS-R2	GCCTATTTGTAGCATTTCTGTCATT GAACATCTCAGTCTAGAATCTGAT
Secuenciación de RIP-3	RIP-3-F RIP-3-R	AATCACACCTCAGGATTTTCAT GATGGGTCATAAGGTTTTGC
Sondas ARN (REMSA)	RIP-3 RIP-3 <sup><i>VRN-A1</i></sup> RIP-3 <sup><i>VRN-D4</i></sup> RIP-3L RIP-3L <sup><i>VRN-A1</i></sup> RIP-3L <sup><i>VRN-D4</i></sup>	GUCAUUGUUGUUGGUAUGGAUCGUAUG-Biotin GUCAUUGUUGUUGGUAUGGACCGUAUG-Biotin GUCAUUGUUGUUGGUAUCGACUGUAUG-Biotin GUCAUUGUUGUUGGUAUGGAUCGUAUGCCAAAAA-Biotin GUCAUUGUUGUUGGUAUGGACCGUAUGCCAAAAA-Biotin GUCAUUGUUGUUGGUAUCGACUGUAUGCCAAAAA-Biotin

\* la letra minúscula indica un mismach introducido para incrementar la especificidad de la reacción de PCR. \*\* Los cebadores para amplificación de ADN copia fueron diseñados previamente (Chen et al. 2012), estos cebadores amplifican simultáneamente *VRN-A1* y *VRN-D4* y son diferenciados mediante digestión con la enzima *Bst*NI.

**Tabla A.4. Lista de los BACs de la biblioteca del cromosoma 5DS secuenciados.** Esta biblioteca fue generada por el Instituto de Botánica Experimental (República Checa) y los BACs son descriptos de acuerdo a su nomenclatura. La secuenciación confirmó que los contigs CTG87 y CTG61 están conectados.

<b>No. en Figura 3.6</b>	<b>Nombre</b>	<b>Contig</b>
<b>1</b>	TaeCsp5DShA_0045_O10	CTG87
<b>2</b>	TaeCsp5DShA_0085_P15	CTG87
<b>3</b>	TaeCsp5DShA_0056_K08	CTG87
<b>4</b>	TaeCsp5DShA_0076_N19	CTG87
<b>5</b>	TaeCsp5DShA_0038_M14	CTG87
<b>6</b>	TaeCsp5DShA_0036_J13	CTG87
<b>7</b>	TaeCsp5DShA_0082_O19	CTG61
<b>8</b>	TaeCsp5DShA_0043_I14	CTG61
<b>9</b>	TaeCsp5DShA_0035_M17	CTG61
<b>10</b>	TaeCsp5DShA_0058_E14	CTG61
<b>11</b>	TaeCsp5DShA_0086_K20	CTG61
<b>12</b>	TaeCsp5DShA_0039_M05	CTG61

**Tabla A.5. Entradas donde la region RIP-3 en VRN-A1 fue secuenciada. Los nombres comenzando con ‘PI’ o ‘Citr’ son identificadores de GRIN, Los numerosos comenzando con ‘UH’ son identificadores de la coleccion de germoplasma de trigo de la Universidad de Haifa (Israel) (Nevo et al. 1989 y Peleg et al. 2005), y ‘MG’ o ‘IDG’ corresponden a identificadores de la coleccion de *T. dicoccum* de la Universidad de Bologna (Italia).**

Trigo	Locación	Identificador
<i>T. urartu</i>	El Beqaa, Lebanon	PI 428275, PI 428280, PI 428282, PI 428283, PI 428285, PI 428286, PI 428287, PI 428288, PI 428291, PI 428292, PI 428294, PI 428295, PI 428296, PI 428304, PI 428321, PI 428327, PI 538741, PI 428279, PI 538740, PI 428293, PI 428276, PI 428222, PI 428223, PI 428240, PI 428244, PI 428251, PI 428239, PI 538733
	Urfa, Turkey	
	Mardin, Turkey	PI 428201, PI 428202, PI 428228
<i>T. dicoccoides</i>	Beir-Oren, Israel	UH 28
	Daliyya, Israel (south Haifa)	UH 29
	Diyarbakir, Turkey	PI 428028, PI 428041
	Gamla, Golan Heights	UH 08
	Iraq	UH 41, UH 42
	J'aba, Mid-north Israel	UH 23
	Kokhav-Hashahar	UH 19
	Mt. Gerizim, Israel	UH 17
	Rosh-Pinna	UH 09
	Seryia	UH 40
	Golan Heights	UH 11
	Yabad, West Bank	UH 32
	Yehudiyia	UH 07
<i>T. dicoccum</i>	Armenia	MG 5274
	Ethiopia	MG 5306
	Georgia	MG 5314, MG 5315
	Germany	MG 5398
	Iran	MG 5273, MG 5275, MG 5276
	Italy	IDG 8634
	Kenia	MG 3428, MG 3430
	Palestine	PI 355496
	Russia	MG 5269, MG 5270
	Turkey	PI 94626
<i>T. aestivum</i>	Canada	Norstar (Citr 17735), Fife (PI 283820)
	Argentina	Barleta (Citr 8398)
	Australia	Falcon (PI 292578), Golden drop (PI 92399)
	Sweden	Kronen (PI 278526)
	United States	Little club (Citr 4066), Sterling (Citr 17859), Finch (PI 628640), Alpowa (PI 566596), Louise (PI 634865), Eltan (PI 536994)

**Tabla A.6. Descripción de las entradas de *T. aestivum* utilizadas para los análisis de diversidad genética.**

Los números corresponden a los utilizados en las figuras 3.16 y 3.17. Los nombres comenzando con ‘PI’ corresponden a la red de germoplasma (GRIN) ([www.ars-grin.gov/npgs](http://www.ars-grin.gov/npgs)) y los nombres comenzando con la letra ‘K’ corresponden a números de catálogo del Instituto Novosibirsk (Rusia). Las entradas de *T. aestivum* spp *aestivum* fueron descritas con anterioridad (Wang et al. 2014). Las últimas dos columnas indican los números de identificación usados en Fig. 3.16B y Figura 3.17.

Identificador	Descripción	Locación	Hábito	Fig. 5B	Fig. S11
PI 272580	sphaerococcum	Pest Hungary	-	1	94
PI 352499	sphaerococcum	India	-	2	96
PI 282451	sphaerococcum	India	-	3	95
PI 330556	sphaerococcum	England United Kingdom	-	4	97
Tandem	aestivum		winter	5	1
Gent	aestivum		winter	6	2
Colt	aestivum		winter	7	3
Chisholm	aestivum		winter	8	5
Custer	aestivum		winter	9	6
OK05526	aestivum		winter	10	4
Bronze	aestivum		winter	11	10
PI 352299	compactum		-	12	13
Antelope	aestivum		-	13	8
OK04505	aestivum		winter	14	7
PI 361807	compactum		-	15	15
PI 352302	compactum		-	16	12
OK05723W	aestivum		winter	17	9
PI 572913	macha		-	18	63
PI 599665	compactum		-	19	64
PI 372151	compactum		-	20	11
PI 352319	compactum		-	21	14
PI 656959	compactum		-	22	21
PI 542466	macha		-	23	61
PI 352466	macha		-	24	60
PI 361768	spelta		-	25	17
PI 355508	macha		-	26	62
PI 278581	compactum		-	27	16
PI 129523	compactum		-	28	22
PI 324491	sphaerococcum	Delhi India	-	29	23
PI 412975	compactum		-	30	19
PI 60740	compactum		-	31	20
Cltr 4528	sphaerococcum	Punjab Pakistan	-	32	81
PI 337997	sphaerococcum	India	-	33	82
Cltr 4530	sphaerococcum	Punjab Pakistan	-	34	72
PI 324492	sphaerococcum	Delhi India	-	35	69

**Tabla A6.** Continuación.

Identificador	Descripción	Locación	Hábito	Fig. 5B	Fig. S11
PI 277141	sphaerococcum	Germany	-	36	79
PI 277165	sphaerococcum	Punjab Pakistan	-	37	80
PI 115818	sphaerococcum	Punjab India	-	38	75
Cltr 9054	sphaerococcum	Iraq	-	39	66
PI 282452	sphaerococcum	India	-	40	68
PI 272581	sphaerococcum	Pest Hungary	-	41	67
Cltr 4529	sphaerococcum	Punjab Pakistan	-	42	74
Cltr 12212	sphaerococcum	Kansas United States	-	43	78
PI 191301	sphaerococcum	Portugal	-	44	77
PI 352498	sphaerococcum	India	-	45	65
K-5498	sphaerococcum		-	46	73
K-23822	sphaerococcum		-	47	70
K-23792	sphaerococcum		-	48	83
PI 40944	sphaerococcum	Punjab Pakistan	-	49	76
PI 182118	sphaerococcum	Sind Pakistan	-	50	71
PI 42013	sphaerococcum	Uttar Pradesh India	-	51	84
Cltr 4924	sphaerococcum	Uttar Pradesh India	-	52	86
PI 277164	sphaerococcum	Punjab Pakistan	-	53	85
PI 572857	compactum		-	54	92
PI 367200	spelta		-	55	87
PI 572851	compactum		-	56	93
PI 367199	spelta		-	57	88
Cltr 8610	sphaerococcum	Jiangsu China	-	58	90
Cltr 10911	sphaerococcum	Shanxi China	-	59	91
PI 262666	compactum		-	60	89
PI 572912	macha		-	61	50
PI 290507	macha		-	62	54
PI 355511	macha		-	63	52
PI 278660	macha		-	64	55
PI 323436	macha		-	65	51
PI 572905	macha		-	66	53
PI 355514	macha		-	67	56
PI 611470	macha		-	68	57
PI 361862	macha		-	69	59
PI 428146	macha		-	70	58

**Tabla A6.** Continuación.

Identificador	Descripción	Locación	Hábito	Fig. 5B	Fig. S11
WA8100	aestivum		Spring	71	29
Tara2002	aestivum		Spring	72	32
Macon	aestivum		Spring	73	34
H0900009	aestivum		Spring	74	33
Alturas	aestivum		Spring	75	30
IDO377s	aestivum		Spring	76	28
PI 277142	sphaerococcum	India	-	77	31
H0800080	aestivum		Spring	78	35
9245	aestivum		Spring	79	25
9248	aestivum		Spring	80	26
PI 190982	sphaerococcum	Belgium	-	81	99
PI 323439	sphaerococcum	Vienna Austria	-	82	98
PI 278650	sphaerococcum	United Kingdom	-	83	100
PI 572850	compactum		-	84	24
Treasure	aestivum		Spring	85	27
PI 25970	compactum		-	86	18
PI 56213	compactum		-	87	49
PI 323438	spelta		-	88	40
PI 295059	spelta		-	89	37
PI 306550	spelta		-	90	36
PI 428178	macha		-	91	39
PI 355642	spelta		-	92	42
PI 631161	spelta		-	93	43
PI 348303	spelta		-	94	41
PI 221419	spelta		-	95	38
PI 192717	spelta		-	96	44
PI 348033	spelta		-	97	45
PI 348273	spelta		-	98	48
PI 378469	spelta		-	99	46
PI 469032	spelta		-	100	47